

7 January 1982

Reagan's mistake on Soviet sanctions

The US Administration, seeking to punish the Soviet Union for what is happening in Poland by suspending an exchange agreement, has revealed its ignorance of what science is for.

Towards the end of the Napoleonic wars, Sir Humphry Davy went on a journey on the European mainland untrammelled by the circumstance that Britain was literally at war with France (and that most other European states were at war with at least one other). The journey, frequently described as an illustration of how even then the international interests of science took precedence over international politics, is made still more pointed by the knowledge that Davy was combining business with pleasure — his honeymoon. Personal travel and national hostilities were, it seems, independent of each other. Now all that has changed. President Ronald Reagan's suspension last week of the exchange agreement between the United States and the Soviet Union, part of his programme of sanctions against the Soviet Union in response to the supposed Soviet support of martial law in Poland, will effectively put latter-day Davys in their place. It is also a stupid and a dangerous move. It should be quietly abandoned.

Programmes of economic and cultural sanctions have been frequently tested in the years since the Second World War, and have as frequently been found wanting. The best thing to be said about them is that they are less damaging than outright warfare. They are also, usually, ineffectual, as President Carter found after Afghanistan. On this occasion, the difference of opinion between some Western European governments and the United States on the wisdom of sanctions as a means of interdiction will further undermine what President Reagan can hope to achieve — which by all accounts (see page 3) is not in any case very much. It is however clear that neither he nor his advisers appreciate the enormity of including within a set of sanctions the suspension of a longstanding programme for the exchange of academic scientists between one country and another.

This is why. Open warfare is a means by which one nation state can legitimately (*pace* the pacifists) enforce its will on another. Even nations that unfashionably declare war on others usually express their hope that in due course all will again be sweetness and light. Programmes of sanctions, damage-limiting though they may be, also anticipate the earlier return of normalcy. So sanctions should be designed to maximize the short-term damage done but to create as few impediments as possible for the hopefully happy years ahead. Denying the Soviet Union the right to buy pipe-laying machines in Texas is unambiguously in the former category, suspending the exchange agreements unfortunately in the latter.

That there should have to be formal schemes for enabling scientists from one country to work, if they wish, in laboratories in another is, of course, in itself a mark of failure. If Dr X should wish to work in Dr Y's laboratory and Dr Y should be agreeable, and if one or other should be able to find the necessary funds, why should governments be involved? People like Mr Frank Carlucci, Deputy Secretary at the Pentagon, think they know why not: if Dr X is a Soviet citizen, he will be part of a "highly orchestrated, centrally directed" effort to pick the brains of unsuspecting American scientists. Mr Carlucci is naturally right to say that Soviet participants in the bilateral exchange agreement benefit from their visits to the United States, and are afterwards potentially more useful as applied scientists, sometimes in military programmes. But is there any evidence that their American hosts, the Dr Ys of this world, are entirely without

benefit? If so, the exchange programme has been maladministered — visitors have been foisted on unwilling hosts. If not, Mr Carlucci should embark on what would be called an evaluation of the benefits and disbenefits of the exchange programme. He would come to two conclusions — first, that the total benefit from mutually sought exchanges for the research enterprise is greater than the sum of the individual benefits and, second, that his government's proper course of action, as a signatory of the Helsinki accords, is to insist that the then agreement, at least in principle, that people should be more or less free to move to where their inclinations take them, should be honoured.

That, unfortunately, will seem too abstract a course to the Reagans and Carluccis of this world. The Soviets *are* benefiting, and must be prevented from doing so, they will say. And we know that Helsinki doesn't work, don't we? That, of course is true, even if the statement of the truth comes oddly from the inheritors of those who negotiated the agreement. But the reasons why Helsinki has failed in respect of scientific exchanges need more careful consideration. The objective now should be to restore some semblance of the *laissez-faire* of Humphry Davy's time. The only way of doing that is to shift responsibility for the administration of exchanges from the State Department (and the Pentagon) to some convenient halfway house such as the National Academy of Sciences. That stratagem has been adopted elsewhere, even in backward Britain. Perhaps the time has come when the National Academy of Sciences in Washington should insist that no other body, not even the State Department, can do what is required. If necessary, it should be prepared to raise the funds required privately, from the foundations and from industry.

More change now due

Reform of the British research council system is overdue. Does the government have time?

The British government has mostly done the decent thing in its dealings with the publicly-supported research councils. Broadly speaking, their budgets for the financial year beginning on 1 April are unchanged (see page 4). So the Prime Minister will be able to tell the British electorate at the next general election, some time before 3 May 1984, that she has "protected" publicly-supported civil science from damaging economies. On present form, the government may have very little else to boast about at the end of what is almost certainly its last full financial year in office. But this is not a vote-catching issue. Thus the government's generosity towards the research councils — a curious description of a recipe for continued stagnation, but appropriate in the circumstances — must be taken as deliberate. Will it therefore, in the remainder of its time in office, put in hand the structural reforms of the mechanisms for supporting the five research councils that are now long overdue?

The research councils are inevitably a mixed bunch. But even those with responsibilities for fostering basic research in fields such as medicine or agriculture are important sources of support for university research. The Science and Engineering Research



Council, on the other hand, although enjoined to support academic research, has for the past decade been flirting with ways of making university research more directly relevant to British industry. On the grounds that all of the councils have a finger in the university pie, their funds are channelled through the Department of Education and Science, but the total sum available each year has for the past decade been divided between the various councils on the recommendation of the Advisory Board for the Research Councils. Even when this mechanism was established, in the wake of the Rothschild reorganization in 1971, it was feared that the board, on which the executive officers of the research councils are necessarily influential, would become an arid device for sharing out funds in such a way as not to give too much offence. So it has turned out.

The advisory board is thus an obvious candidate for change. But how? There is much to be said for the recipe offered last November by the House of Lords Select Committee on Science and Technology — appoint a full-time chairman, give him (or her) a small staff and require the board to comment on the policy questions that now go largely by default. Under such an arrangement, the chairman of the board would become what the Department of Education and Science now conspicuously lacks — a chief scientist of a kind. These proposals have the advantage that they could be put into effect without much trouble. They would also be improvements. The government should listen to what the House of Lords has said, and quickly.

There are, however, snags. The reform of the Advisory Board for the Research Councils along the lines suggested by the House of Lords committee would leave untouched the habitual secrecy of advisory bodies working for British government departments. The theory is simple; advisory bodies exist to give advice which ministers may, at their pleasure, decline. And, the argument goes, it would be unseemly, cramping and politically irksome if the rejection by a minister of an advisory committee's advice were publicly advertised. But this is precisely what has happened with this year's advice on the research council budgets. The decision that the budget of the Social Science Research Council should be reduced by £1.1 million (just over 5 per cent) was taken by Sir Keith Joseph, the Secretary of State for Education and Science, in flat contradiction of the recommendation of the Advisory Board for the Research Councils. No explanation has been given, so that there is no way of telling whether the minister's fiat reflects the fashionable animus against the social sciences or whether there were — as there might well have been — more substantial grounds for his decision. Whatever the truth, the minister is likely to be more seriously damaged by the gossip that will now flourish than he would have been by an open discussion of his reasons. Secrecy, it seems, hurts even those whom it is intended to protect.

The case of the Social Science Research Council illustrates another weakness of the present system. The council was set up nearly twenty years ago to foster academic research in the social sciences. Throughout that period, it has struggled hard to establish academic centres of excellence in the face of open hostility from its enemies and, more disarming, the wayward kookiness of some of its friends. On the grounds of youth, it was exempted from Lord Rothschild's recommendation in 1971 that government departments should pay for the cost of applied research supported by the research councils. Since then, the Social Science Research Council has been complaining that government departments are too fond of stealing its clothes by commissioning basic research from their own resources. Now, ironically, there is to be another inquiry by Lord Rothschild into the application of the customer-contractor principle in this field; one result may be to reconcile the advisory board with its minister.

What will remain undecided are the questions that the advisory board should have had the wit to take up several years ago — is a research council the best mechanism for fostering the social sciences? what criteria of excellence make sense? and what, in any case, is to be made of the peculiarly British sneer that what are called the social sciences are not science? It is easy to understand why the advisory board would have shrunk from such an investigation; if the precedent were established that the constitution of

the newest research council might be questioned, who could then ensure that the well-established councils were free from scrutiny? The advisory board, in other words, is undermined by backscratching which is then strengthened by the convention of secrecy.

The habit of secrecy is not merely damaging to all those concerned but also inappropriate. Over the years, the research councils have (to their credit) established a remarkable sense of trust with their customers in the universities. In Britain as elsewhere, the peer review of research proposals has created a sense that substantial parts of their budgets are spent by the scientific community, not by career officials. Some of the councils have also developed a knack of putting controversial ideas into circulation before they are finally decided. But the questions that the advisory board should be tackling — the place of social science, the health (or sickness) of the dual support system, the success (or otherwise) of the University Grants Committee's designated courses in engineering education and so on — are also questions that cannot be settled convincingly by a small group of people, however appointed.

The lessons for the government are clear. The advisory board should be strengthened along the lines suggested by the House of Lords and, in the process, the dominant influence of the research councils should be deliberately weakened but not abolished. Then, some device must be found for making the advisory board a more public body, able at least to advertise the issues on its agenda and ideally compelled to publish not less than once a year a reasoned account of what it has been doing. There is no substantial danger that in the process the British Constitution would be undermined or the freedom of ministers compromised. On the contrary, ministers would be less at risk than at present of having to make up their minds without the benefit of even half-baked advice.

Not president perpetual

Dr Philip Handler, who died last week served the US National Academy of Sciences well.

Being President of the United States is a relatively simple job; power and patronage help to bring to heel unwilling cabinet colleagues and even members of the Congress. Presidents of the National Academy of Sciences in Washington enjoy no such benefits. Even though elected, they must continually reassert their authority among their electors, many of whom declare that they could do the job just as well if only they chose to spend their time in such a demeaning way. One measure of Dr Philip Handler's achievement in this office is that during his twelve-year stint at the academy, which ended last June, he was twice re-elected.

A large part of the explanation is that he was an articulate man; he could make fine speeches but also talk his way out of trouble. He was also courageous. He fought the extremists among the environmentalists when many others thought it prudent to keep their heads down. He fought Congress on several occasions, not only about the scale of support for research but over occasional threats of interference. Handler's early years at the academy saw the rapid growth of its secular activity — producing reports on topical issues, sometimes important and sometimes trivial — whereafter he had to fight the committees for the right to ensure that their reports were properly reviewed before publication. Not all his battles were successes. The campaign to finance the purchase of Einstein's statue now in the academy's front garden might have been successful if the statue had been better.

Inevitably, continual battles take a toll. Handler, always an iconoclast, occasionally seemed to be oversensitive to criticism. Latterly, he was concerned to know what he would do on leaving the academy and, more recently, was embrittled by his largely secret knowledge of his mortal illness. He will, however, be remembered well, not merely for his wit and eloquence but for having rescued the academy from its earlier spells of idiosyncratic and then indolent leadership — and for ensuring that it is no longer simply an academy.

Anxiety about Reagan's sanctions

Two views on the wisdom of suspension

Washington

President Reagan's announcement last week that the United States is suspending its scientific exchanges with the Soviet Union, as part of a package of reprisals aimed at chastising the Soviet Union for its support of the military clampdown in Poland, has provoked concern in the scientific community that the move indicates an increasing willingness to use science for explicit political purposes.

The immediate effect of the President's decision will be to suspend the scientific exchange agreement signed by President Richard Nixon and Soviet President Leonid Brezhnev in 1972. Furthermore, Mr Reagan announced that unless the situation in Poland changes the agreement will not be renewed when it runs out in July.

Similarly, agreements for cooperation in energy and space research will not, at present, be renegotiated when they run out in May. And a number of other agreements which include scientific components are also being suspended and reassessed.

To a large extent, however, Mr Reagan's decision to suspend scientific exchanges is viewed as more symbolic than substantive. In the case of high-technology trade, an area announced for suspension, even the US Department of Commerce accepts that the Soviet Union is little more than a "marginal market" for sophisticated equipment, accounting for less than one per cent of the US electronics industry's annual sales of \$200,000 million, and that the Soviets may well be able to meet their needs from other sources.

Officials at the National Science Foundation, which is responsible for nine of the twelve working groups still operating under the terms of the 1972 agreement, argue that as a result of deliberate efforts over the past few years to eliminate exchanges involving a marked imbalance of benefits, those still in effect provide scientific benefit to both sides.

Many — although not all — US scientists fear that if even these exchanges are now halted, the effect could be counter-productive. Not only would it close off an important channel of communication between two intellectual communities, but it would provide a useful propaganda weapon for use against those attempting to separate politics from science.

"I feel that it is very important that these communication channels are kept open because they provide a crucial link between

different societies and different peoples during periods of difficulty, and because they can be kept separate from immediate political concerns", Dr Alan Bromley, professor of physics at Yale University and president of the American Association for the Advancement of Science, said on the eve of the association's annual meeting.

At the National Academy of Sciences, which two years ago suspended all bilateral symposia, seminars and workshops with the Soviet Academy of Sciences in protest at the treatment of physicist Dr Andrei Sakharov, but has made no move to interrupt the exchange of individual scientists between the two academies, president Dr Frank Press said he was concerned that if there was no formal scientific agreement at all between the United States and the Soviet Union, it would be difficult to keep channels of communication open.

Even some of the groups which have been most vocal in their protests over the treatment of Dr Sakharov and other dissident scientists have expressed fears that their effectiveness could be severely reduced if scientific exchanges were terminated. Dr Max Gottesman, a member of the Committee of Concerned Scientists, said last week that since Soviet diplomats did not seem to consider the termination of such exchanges to be a major blow to their domestic scientific effort, the United States stood to lose more than it might gain in terms of influence over opinion in the Soviet scientific community.

In contrast, there is less consensus on how far the federal government should go in restricting access by foreign scientists, in particular those from the Soviet Union, to

US research laboratories on the grounds that the knowledge they pick up might later be used for military purposes.

Following protests from the scientific community that such restrictions could unnecessarily restrict the flow of scientific information, Mr Frank Carlucci, deputy secretary of the US Defense Department, has provided a detailed list of instances where, he claims, the Soviet Union has achieved significant military or technical benefits through visits to US laboratories made by Soviet scientists on exchange programmes. According to Mr Carlucci, "it is quite apparent the Soviets exploit scientific exchanges as well as a variety of other means in a highly orchestrated, centrally-directed effort aimed at gathering the technical information required to enhance their military posture".

A significant number of scientists are known to be in general agreement with the Defense Department's position. Dr Bromley, for example, said that although he felt restrictions should not be applied to basic science, he was concerned that information of potential military value might have been "given away" through exchange programmes.

The National Academy of Sciences is considering setting up a committee to look at the implications for the scientific community of existing and proposed federal restrictions on the flow of technical data. Dr Richard D. DeLauer, Under-Secretary of Defense for Research and Engineering, is said to be in favour of such a committee, although academy officials fear the government might act before a proper study can be carried out. **David Dickson**

Polish academia only partly normal

Polish academic life formally resumed this week but was, however, only partial. There are reliable reports that only graduate and fifth-year (finals) students have been readmitted to the universities and that all other undergraduate courses have been suspended until next year.

This leaves in doubt the status of the suspended students. A decree of the Military Council, passed on 30 December, obliges all able-bodied males between 18 and 45 years of age to work for the duration of martial law. Students as such are exempt, but it is not clear whether the exemption applies to those who must wait until next autumn to resume their studies.

There are also unconfirmed reports of some first-year students being drafted for military service. University graduates who have not yet found work are obliged to report to plenipotentiaries from the Ministry of Labour, Wages and Social Affairs stationed at their former colleges. Since unemployment among young graduates is considerable — particularly in the humanities and social sciences — there

seems little chance that for them the clause in the decree, enjoining the administrators to take professional qualifications into consideration "as far as possible", will be of more than a token significance.

Academic staffs are far from happy with the new situation in Poland. On 23 December, the Prime Minister, General Wojciech Jaruzelski, together with Deputy Prime Minister Mieczyslaw Rakowski and Politburo member Hieronim Kubiak, met 69 leading academics to discuss what Jaruzelski described as "the active role of Polish scientists, intellectuals, writers and artists working for the salvation of the homeland, consolidation of the state and the building of a bridge of special patriotic agreement". Both Kubiak and Rakowski have the reputation of being liberals, and the academics seem to have been selected mainly from those who played little or no active part in the campaign for academic autonomy of the past 16 months.

Even so, no agreement was reached. The official communiqués spoke merely of a "long and frank discussion", with the

academics stressing the "urgent need to rebuild trust and calm scientific teaching in Polish universities". According to an unofficial *Solidarity Information Bulletin* now circulating in Warsaw, the seven-hour meeting was deadlocked by the persistent demand of the academics for the release of all detainees and for the participation of the university rectors recently elected under the new democratic procedures in any further negotiations.

The presence at the meeting of Dr Aleksander Gieysztor, president of the Academy of Sciences, must have sharpened the discussion of the detainees. He was himself picked up in one of the early round-ups, although released shortly afterwards.

The unofficial lists of detainees now being compiled reflect just how active a role the academic community played in the *Solidarity* movement. They include:

- Lecturers from the unofficial Society of Academic Courses (the "Flying University") Stefan Amsterdamski, Władysław Bartoszewski, Andrzej Celinski, Jerzy Jedlicki and Jan Walc.

- Solidarity advisers from the universities, including Dr Bronisław Geremek, a Warsaw lecturer in economics, who was a member of the *Solidarity* delegation to France last October, and Dr Stefan Kurowski who drew up the project for economic renewal discussed at the *Solidarity* Congress.

- Doctors and medical workers including Grazyna Przybylska-Wendt from the *Solidarity* presidium.

- Academy or university employees who, until now, have never achieved this type of prominence — among whom Dr Stanisław Kozłowski, the new rector of Poznań University, is the most eminent so far.

The military rulers seem willing to make some small concessions to the academics — the constitution of the Polish People's Republic (Art.5.3) enjoins them to maintain the "constant progress of science and technology". Academic detainees from the Warsaw area are housed in a government hotel near Drawsko. Earlier reports that *Solidarity* advisers Jacek Kuron and Adam Michnik had been savagely beaten are unfounded.

The main instigators of the academic turmoil, according to the official statements, were the activists of the Independent Students Union (NZS). These, it is said, had tried to turn NZS into a political organization, rejecting the policy of social accord and "stirring up anti-Soviet sentiments" — a line which suggests that a purge of the student body may be forthcoming.

So far, no mention has been made of the fact that, only a few days before the military take-over, the Conference of University Rectors, meeting in Poznań, had formally endorsed the recent NZS protests as a valuable contribution to the struggle for academic freedom.

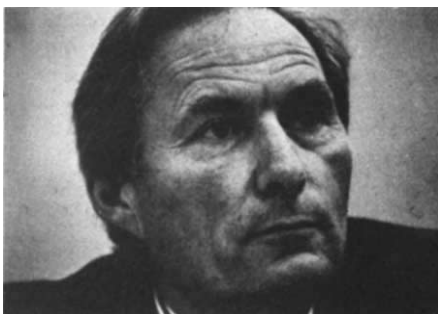
Vera Rich

High-energy physics

All change

With approval for its latest accelerator, LEP, under its belt, the European Centre for Nuclear Physics, CERN near Geneva, seems prepared to experiment — not so much with particles but with its governing body. For alongside the announcement of the CERN budget, and the final approval for LEP (see *Nature* 24/31 December, p.685), comes news of the appointment of two active new characters to the CERN Council. Sir Alec Merrison, currently chairman of the UK Advisory Board for the Research Councils and vice-chancellor of the University of Bristol, is to become president of the council; and Umberto Vattani, a high-ranking Italian civil servant who was involved in a disagreement two years ago between Italy and Germany over the appointment of a German (Professor Herwig Schopper) as director-general of CERN, becomes one of the two vice-presidents.

Sir Alec was himself a high-energy physicist, and in fact in 1959 he was co-author of the first experimental paper to



Merrison: chairman everywhere?

come out of CERN. He was also a director of the now defunct Daresbury electron accelerator, NINA, and has defended the place of "big science" in the highly-constrained UK science budget.

On his appointment — which gives him control of the CERN Council's committee and subcommittee work during the next year, as well as presidency at the council itself — Sir Alec said he was delighted at the decision over LEP. "All the delegates are amazed that CERN has been able to pull this thing off," he said, but at the same time the next few years would see CERN taking "a jolly dangerous path".

This is because LEP is to be built out of the current budget only by making massive savings in the present programme. Closing certain facilities entirely would save something, but most of the savings will have to come from running down the experimental work on the CERN 400-GeV super proton synchrotron, the laboratory's biggest machine. There will therefore be a hiatus in research.

Vattani, on the other hand, now reassured that Schopper has proved an "excellent" director, and that Germany will not try to squeeze CERN in favour of

its own national accelerators, sees looming problems in personnel and pensions

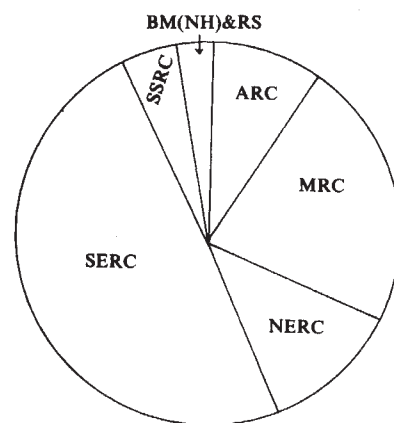
Other questions which will arise for the council in 1982 are late payment (some members have indicated that they may not be able to pay their subscriptions on time) and the management of the LEP project itself. LEP will span two countries, be 27 km in circumference, cost £275 million, and will still face some environmentalist opposition in the region. Professor Schopper will no doubt welcome all the help that the council can give him in this, the largest ever CERN project.

Robert Walgate

Hardly any change

The arguments put forward by the past and present chairmen of the British Science and Engineering Research Council (SERC) that the council should take a larger slice of the science vote have modestly borne fruit. Although the civil science budget for 1982-83, made public a few days before Christmas, has roughly kept its real value, SERC's budget has been increased by about 1 per cent. Most of the council's extra money will be spent on biotechnology, information technology and on helping universities to maintain the quality of research in spite of cuts in their own budgets.

It is equally no surprise that the council to come off worst in this year's division of the science vote is the Social Science Research Council (SSRC), with a real budget for 1982-83 about 4 per cent less than in 1981-82. SSRC has suffered substantial cuts in the real value of its grant for the past three years, reflecting uncertainties about the role of social science research in Britain.



The three other research councils, however, are to receive the same real budget in 1982-83 as in 1981-82. All, however, will be concerned that the government has allowed for only a 4 per cent increase in salaries over the year. If, as last year, the increase in the salary bill is higher, then the amount of research the councils can buy is bound to decrease.

Judy Redfearn

American science

Supply-siders now

Washington

"We are now forging a substantially closer relationship with both the legislative and executive branches of Congress", the retiring president of the American Association for the Advancement of Science (AAAS), Dr Alan Bromley of Yale University, said on the opening day of the association's annual meeting in Washington on Monday.

Dr Bromley listed various ways in which AAAS was exercising what he described as a "service role" for private and public science decision-makers. Such activities range from providing testimony to congressional hearings to preparing regular and widely quoted assessments of the impact of federal policy on the size of the US research and development effort.

But if a considerable amount of AAAS's time is now spent helping the government to do its job, there has also been an expansion of effort to compensate for federal actions — for example, the cutting back of support for science education in schools.

Following up an initiative launched at last year's annual meeting in Toronto, the chairman of the AAAS board of directors,

Dr Frederick Mosteller, announced that the association had just received a \$6.5 million grant from Phillips Petroleum for a series of television films to complement the work of mathematics teachers in public schools "by making mathematics more lively and significant".

Another educational initiative is the publication of a set of science teacher notes based on articles appearing in AAAS's monthly science magazine *Science 82*. The association is also convening two conferences later this year with leaders of various scientific and technical societies, the first dealing with urgent policy issues in science education, the second concentrating on the growing shortage of science and mathematics teachers.

Inevitably, a prominent theme of this year's meeting is the impact of the Reagan Administration's cuts in public spending on scientific research. In an address on Sunday evening, the President's Science Advisor, Dr George (Jay) Keyworth, emphasized again the Administration's awareness of the importance of basic science, promising that the next "budget submission will clearly reflect that".

He also again warned that there were hard choices to be made. In the short term, there was general acceptance at the meeting that such a policy was appropriate. Dr Bromley pointed to the Department of

Energy's High Energy Physics Advisory Board and the Nuclear Science Advisory Committee as examples of ways that scientists in a particular fields, when finances get tight, have been able to determine their own scientific priorities.

But in the longer term, it was said, repeated budget cuts will be severely disruptive. Figures produced by AAAS itself last week showed that for the two-year period 1981–82, the result of federal actions will be to reduce the civilian research and development budget by about 16 per cent in real terms (constant dollars), with a 22.2 per cent growth in the military research and development budget over this period.

Another issue being hotly debated in the meeting rooms and corridors is the current dispute about creationism and the teaching of evolutionary theory in public schools. On Monday morning, the AAAS board unanimously passed a resolution that "creationist groups are imposing beliefs disguised as science upon teachers and students to the detriment and distortion of public education in the United States".

Registration at the 1982 meeting is expected to be about 5,000 which AAAS officials say is approximately the same as the last time the association held its annual meeting in Washington three years ago. Next year, in Detroit, they are hoping for something better, by moving the timing of the meeting from January to May.

That decision has been partly determined by the weather. During last year's meeting, Toronto experienced its coldest winters for almost a century, with temperatures of -30°C . The previous year, San Francisco airport had been closed by fog.

David Dickson

UK director for US Biogen laboratory

Dr Richard Flavell, one of Britain's whizz-kids of gene cloning, is to leave the country to become research director of Biogen Inc., the American arm of the international biotechnology firm, Biogen NV.

Flavell is currently leader of the only gene cloning group at the National Institute of Medical Research (NIMR), Mill Hill, London, one of the two principal laboratories of the UK Medical Research Council. Gene cloning is also done at other laboratories of the council — notably at Hills Road, Cambridge — but there is a strong feeling that application of the technique, central now to the whole of molecular biology, should be more in Britain. The loss of Dr Flavell is thus significant, particularly at a time when NIMR is seeking a new director and re-considering its role within British biology.

In fact, Flavell has officially been granted "leave of absence" from Mill Hill for two years; but it seems unlikely that two years hence Flavell will want to leave Cambridge, Massachusetts, where the Biogen laboratory is being set up.

Flavell will continue to direct a group at Mill Hill, if at a distance, but he wishes to take some of the more fashionable work with him. He expects that half-a-dozen of his Mill Hill team will go with him to Biogen, taking with

them expertise on the H-2 regions of the mouse genome. These regions code for histocompatibility antigens in the mouse, and it may be speculated that Biogen will now take an interest in the generation of reagents for tissue-typing in man, using the equivalent human HLA regions. Flavell will leave at Mill Hill the part of his group which deals with the globin genes, also important in immunology but less at the centre of attention now than H-2 and HLA.

At Biogen, Flavell will take over a fledgling staff of 20 scientists, which he will increase to 75 by the end of 1982. He will, of course, be concerned mostly with fostering the commercial interests of Biogen through research, but has been guaranteed the opportunity to pursue his own basic research — largely through the Mill Hill team he takes with him.

Flavell cannot say yet what the research interests of his new laboratory will be; but according to Professor Charles Weissmann of Zurich, who is chairman of the scientific council of Biogen, the Cambridge laboratory was originally planned to pay closer attention to the chemical and processing aspects of biotechnology than Biogen's Geneva laboratory which was founded earlier (see *Nature* 27 October 1981, p.599). Whether Flavell will in fact follow this line is open to question.

Robert Walgate

Biomedical research

NIH takes a cut

Washington

The National Institutes of Health (NIH) told recipients of biomedical research grants and contracts last week that new and continuation grant awards made in the next three months are likely to be 4 per cent lower than initially proposed by President Reagan in his budget request to Congress last March.

It could have been worse. A revised budget request, issued in September, demanded a further 12 per cent reduction in all discretionary public spending and many federal agencies have had their activities curtailed significantly in the drive to reduce the federal deficit. As it is the NIH budget for 1982 — the fiscal year which started last October — will only increase by 2.2 per cent, considerably less than the current rate of inflation, and the first significant drop in real terms for federal support of biomedical research since the mid-1970s.

However, in the short term, no major damage seems to have been inflicted, and

NIH's budget strategy of attempting to stabilize the number of new and renewing competitive awards at about 5,000 remains virtually intact. Furthermore, Congress has replaced much of the money for research training grants which the Office of Management and Budget wished to cut.

Technically, the 4 per cent reduction is only a provisional figure, valid up to March. Unable to agree on a detailed appropriations bill, Congress agreed last month to a "continuing resolution" for the next three months. In practice, however, since debate on the budget for the 1983 fiscal year begins later in January — and no substantial objection to the provisional 1982 totals are expected from the White House — the 4 per cent reduction is likely to run to September.

It has been a confused budget year for NIH. Before leaving office in January, President Carter proposed a biomedical budget for the institutes of \$3,770 million, a figure left virtually untouched in the Reagan Administration's budget review in March and representing an increase of about 5.6 per cent over 1981. However, realizing that the budget deficit in general was going to be much larger than initially anticipated, the Reagan Administration in September recommended an across-the-board cut of 12 per cent, which would have reduced the NIH budget to \$3,310 million.

The net result of last month's continuing

resolution is that, rather than having to reduce the number of new awards for competing research projects to 4,230, as the proposed September revisions would have required, NIH expects to be able to make almost 4,800 of these during the current financial year. This is little short of the 5,000 which was adopted as a goal two years ago, following several years in which the number of such awards, which are provided to support investigator-initiated research projects, had fluctuated widely.

The budget resolution procedure thwarted a move supported by several legislators in the House of Representatives to direct additional research funds to three institutes: the National Institute of General Medical Sciences, the National Institute of Allergy and Infectious Diseases, and the National Institute of Child Health and Human Development. These were felt to be suffering more than the others from the squeeze on research funds, in terms of the quality of the research projects that were being refused funding.

NIH officials are optimistic that the 1983 budget which Mr Reagan submits to Congress next month will again contain a small increase for NIH. Secretary of Health Mr Richard Schweiker is said to have argued successfully against major cuts in biomedical research that had been proposed by the Office of Management and Budget.

David Dickson

Molecular biology

Trimming costs

Molecular biologists who are used to relying on the European Molecular Biology Organisation for providing fellowships and other useful means of mobility within Europe may have to pay closer attention to the politics of its budget next December — for the 1982 budget has seen a 3 per cent fall (in real terms) for the second year running.

The secretary of the organization, Dr John Tooze, says that management costs have been trimmed to the absolute minimum, and that for 1982 cuts will have to be undertaken in the EMBO programme itself. The 1982 budget will be DM 3.5 million (£825,000), numerically 8 per cent above last year but 3 per cent below EMBO's calculated inflation figure.

The EMBO scheme accounts for 400 fellowships in any year, plus 40 courses and workshops and a small lectureship programme — which often takes prominent American and other visiting scientists around Europe on a lecture tour. It is this lecture programme which is likely to be cut this year.

If the budget were to be cut again for a third year, the fellowships and workshops would be hit, says Tooze. Then the smaller countries (of the 17 member states of EMBO) would certainly object. These countries are probably the main beneficiaries of the programme, and they in fact constitute a two-thirds majority of states on the council which votes the budget of EMBO each year. Decisions can be taken by such a majority but the remaining states — Britain, France, Germany and Italy — contribute nearly 60 per cent of the budget. The danger is that if the small countries kick up a fuss, the big four could refuse to pay. This year a compromise has been reached. It may not be so easy next year.

Robert Walgate

Microbial collections

Bugs in hazard

Once again, a British microbial culture collection is threatened by lack of funds. At the end of last year the Ministry of Agriculture, Fisheries and Food announced its intention to withdraw support for the National Collection of Industrial Bacteria housed at the Torry Research Station in Aberdeen. This is the third time in two years that culture collections have been threatened by withdrawal of funds. The Brewery Research Foundation and the agriculture ministry have already stopped supporting the national yeast culture collection and the Commonwealth Mycological Institute and Natural Environment Research Council have said that they wish to cut support for the national fungi collection.

Although the problems of the yeast and fungal collections have been at least

New director chosen for NIH

Washington

Dr James B. Wyngaarden, chairman of the department of medicine at Duke University in North Carolina, is expected to be nominated by President Ronald Reagan as the new director of the US National Institutes of Health (NIH) in Bethesda, Maryland.

Although yet to be formally announced — and subject to approval by the US Senate — Dr Wyngaarden's appointment is thought to have been unofficially approved by the White House, and the delay in making an announcement is the result of the lengthy security and financial checks which all presidential appointees undergo.

Dr Wyngaarden would probably be a popular choice at NIH's headquarters, and with the biomedical research community in general. Those who know him say they expect him to provide strong leadership and vigorously to defend the \$3,600 million research budget of NIH, the principal source of federal funding for biomedical research in the United States.

Dr Wyngaarden has been a close professional colleague of Dr Donald Fredrickson, who resigned as NIH director last July after six years in the post. Together with Dr John B. Stanbury, the three were co-editors of an influential collection of essays, *The*

Metabolic Basis of Human Diseases.

First published in 1960, the book is about to go into its fifth edition, and is considered by many to be a classic in its field.

Dr Fredrickson said last week that he had "no doubt" that Dr Wyngaarden would fill the NIH position extremely well. "He is a good physician, an excellent scientist, and understands the institutional aspects of medical research very well; I have every confidence that he will be an excellent director".

A graduate of the University of Michigan, Dr Wyngaarden is no stranger to Washington science policy circles. He was a consultant to the Office of Science and Technology Policy between 1966 and 1972, and a member of the President's Science Advisory Committee from 1972 until its abolition in 1973. Before his appointment at Duke, he was professor of medicine and chairman of the department at the University of Pennsylvania from 1965 to 1967.

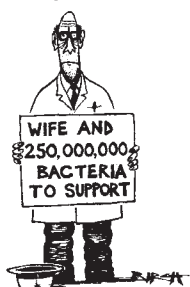
NIH scientists hope that a speedy confirmation of Dr Wyngaarden's appointment will help to accelerate other senior appointments at NIH. Six of the twelve institutes lack a permanent director, and approving the new appointments will be one of the first tasks awaiting the new director.

David Dickson

temporarily solved, the threat to the industrial bacteria collections has stimulated the UK Federation of Culture Collections to urge that a single government agency should take over responsibility for all of Britain's ten national collections. The problem is that the collections currently share seven sponsors, not all of which ascribe high priority to their sponsorship.

In the event, the industrial bacterial collection may also be saved, though not necessarily to the satisfaction of its users. The most likely outcome is that the agriculture ministry, the collection's sole sponsor, will relinquish responsibility to the University of Aberdeen from next April. Negotiations are under way for transferring the collection to the university's company, which handles the commercialization of research relevant to industry. The idea is that the collection should eventually operate as a commercial concern.

Under the arrangements now being discussed the ministry would pay the university for running the collection, although payment would be considerably less than current running costs of about £200,000 a year. Savings would be achieved largely by reducing the number of staff from twelve to about seven. Still to be determined are the university's commission on continuing business, the ministry's initial grant and what should be charged for consultancy and services. Although the proposal would save the collection, the microbiological community, which fears that commercialization could jeopardize some of the services currently on offer and increase costs to users, is glum about the prospects.



Last year, an inter-research council committee was set up to deal with the funding problems of culture collections as and when they arise. But the UK Federation of Culture Collections now argues that the committee's remit is too narrow. Together with three learned societies, including the Society for General Microbiology, the federation is hoping to persuade one government agency to take over responsibility for all ten microbial collections and to coordinate the formulation of a national policy for genome conservation including plant and animal cell lines as well as microbial cultures. As yet, however, no obvious agency has come forward, although the federation is hopeful that a meeting with the Department of Industry early in the year will be fruitful.

Judy Redfearn

Stanford University

At the helm

Palo Alto

When biologist Donald Kennedy became Stanford University's eighth president in August 1980, he pledged strong support for the humanities. Now, more than a year later, he has delivered; \$5.6 million in gifts and grants have helped to create a new Humanities Center. From other contributions, eight professorships have been endowed.

Nevertheless, some faculty members remain sceptical of Kennedy's efforts to make support for the arts and humanities comparable with that for the sciences. One humanist complains that too much time is spent on training and not enough on students' personal development. He thinks that a university should be primarily concerned with values. He is waiting for Kennedy to enunciate his values and to exert intellectual leadership. "You can buy administrative ability", the disgruntled professor believes.

During Kennedy's spell in Washington as Commissioner of the Food and Drug Administration, he won wide respect for his administrative ability. Although he denies rumours about his own political ambitions, he still travels frequently to the capital.

He says that one of his objectives is to improve government understanding of the kinds of problems research universities face and about national science policy. He has the experience and the contacts to talk about these matters in the right places.

And lean, six-foot-one Kennedy believes that Stanford will be able to continue as a first-rate centre of basic research in spite of federal research cutbacks this year of 2 or 3 per cent. He is a self-avowed "relentless optimist". Nevertheless, he worries about reduced support for student financial aid. More than half of Stanford's 12,800 students receive scholarships. Research and training should go hand in hand, a truth not always recognized by the Reagan Administration, Kennedy says.

Kennedy remains doubtful that private support for universities will help to cover the gap created by diminished federal funding. Although private support for research has grown, it amounts to only 3.5 per cent of Stanford's total research budget of \$130 million a year.

The success of Stanford's own venture in commerce is necessarily still unclear. Last month, 71 companies signed licences for the gene-splicing and cloning patents developed by Stanford geneticist Dr Stanley Cohen and University of California biochemist Dr Herbert Boyer. Both scientists have made over their share of royalties to their universities. Initially, the licensing will yield \$710,000 annually, but the long-term potential could be much higher.

In the face of his campus's burgeoning



Donald Kennedy keeps ahead

biotechnology, Kennedy spends much of his time on the ethical questions of proprietary research. He has organized a conference in March with administrators, faculty scientists and representatives of industry to establish possible guidelines for the future. His own interest in science policy stems from his work as a founder of the human biology programme at Stanford. He was chairman of the biology department in 1965-72. His colleagues consider him a brilliant administrator and a fine scientist, able to articulate his views clearly.

Although a member of the National Academy of Sciences, Kennedy does not expect to return to the laboratory. His speciality is the neurophysiology of behaviour. He continues to work on two unpublished papers about how connections among nerve cells are responsible for fixed behavioural acts.

Meanwhile, he enjoys his contact with students and giving guest lectures, as well as advising eight freshmen. He has a reputation as an outstanding and entertaining lecturer. Thus, to illustrate man's evolution from the primitive fish *Amphioxus*, he will sing to the tune of "It's a long way to Tipperary",

Goodbye fins and gill slits,
Hello teeth and hair.

It's a long way from *Amphioxus*,
But we came from there.

Kennedy is blessed with enormous energy. But some of his faculty grumble that he is too enthusiastic, that he is also too generous, and that he resorts too quickly to resolving disagreements with university funds.

Students nevertheless like him. Graduating seniors have repeatedly asked him to address them. He attends more than 24 sherry parties annually in student dormitories.

Will his present enjoyment of his leadership of Stanford last? A close associate suggests that within five years Kennedy will start to think about a new post. It takes that long for the excitement and pleasure of the office to begin to dissipate.

At present it is possible to find Stanford's president at 6.30 a.m. each Tuesday and Friday morning running four miles over hilly terrain. Those who wish to discuss problems or to offer suggestions are welcome to join him. Most of his colleagues agree that one has to be pretty swift to outrun him.

Charlotte K. Beyers

CORRESPONDENCE

The ultimate question

SIR — G.K. O'Neill's statement that we are probably alone in our galaxy (*Nature* 294, 25; 1981) is based on two different arguments — one about the unlikelihood of the phenomenon of life and the second about the probability of the development of space colonies by any long-lived civilization. I show here that there is an unexpected link between the two arguments.

It may well be true that the survival of life on Earth is due to an extraordinarily lucky series of accidents by which the surface temperature remained remarkably constant. On the other hand, J.E. Lovelock's suggestion that life itself was and is responsible for the maintenance of a life-friendly temperature, is not without evidence. In his book *Gaia, a New Look at Life on Earth*, Lovelock wrote that the fact that the Earth's climate has changed very little since life appeared, despite variations in the output of heat from the Sun, was one of the reasons for developing the Gaia hypothesis. This hypothesis proposes that "the entire range of living matter on Earth, from whales to viruses, and from oaks to algae, could be regarded as constituting a single living entity, capable of manipulating the Earth's atmosphere to suit its overall needs and endowed with faculties and powers far beyond those of its constituent parts". In another version of the Gaia-hypothesis Lovelock calls this entity an organism, but he prefers to talk about the "Gaia system". The word "organism" is indeed confusing when used to refer to an entity of which the constituent parts are organisms. But I see no reason why this "Earth life system" (having assumed its existence) shouldn't have the property that distinguishes living systems from dead matter, namely the capability to procreate.

O'Neill's second argument — "The development of space colonies . . . is a natural and not particularly difficult stage in the development of any intelligent civilization" — can be rephrased, if we are prepared to admit that the human species and its technology is integrated in the whole of living nature, to read: "The development of space craft is a natural process in every planetary life system and will lead to the creation of independent descendants (space colonies), so to ensure the pre-existence of the planetary life system".

So — is there an association between the hypothesis of an overall "organism" and the (also hypothetical) development of space colonies? It is a question that goes beyond science, and gives us a choice between heroic anthropocentrism and a more objective but difficult to accept biocentrism.

ROELAND A. DE BIE

Rotterdam, The Netherlands

Base mettle

SIR — I was pleased to note that your editorial staff has at last discovered the greatest of American sports as reported in your column "Shame on New York" (*Nature* 5 November 1981, p.2). While your comments on the competition " . . . quaintly known as the 'World Series'" were touching and amusing, I was surprised to note a large number of errors of fact in the column. Given *Nature's* reputation

as a scientific journal, I am sure you will want to waste no time in correcting these.

You suggested that good players are able to hit one fair ball in three. Actually, good players will hit about 8 fair balls in 10, with about 3 in 10 resulting in safely reaching the first base. A player who could hit only one fair ball in three would shortly be sent packing for another sport where his talents might be more appropriate to the level of play (cricket perhaps?).

You noted that the Yankees were not as skilled in catching and throwing the ball as the Dodgers. This is a strange comment on an event in which one member of the Dodgers set the all-time record for most errors (serious mistakes in catching or throwing) in the "World Series". One might also have noted that members of the New York Yankees Baseball Club, once "on base", displayed a remarkable propensity for running directly towards the Dodger player in possession of the ball. The only biological parallel I can think for this behaviour is that of certain species of whales running themselves aground and eventually perishing of their own weight. In the interests of science, *Nature* might humbly have suggested that representative members of the Yankees should be selected for inner ear examinations or behavioural analysis during the winter. I believe that Mr Steinbrenner has made similar suggestions, couched in rather different language.

Finally, *Nature's* comments on the random flight of the batted ball are completely inaccurate. Every baseball player is aware that while the flight of any individual batted ball is unpredictable, when a large sample of batted balls from any one individual is examined, certain patterns emerge. As more and more samples are taken, a pattern of probabilities (not unlike probability patterns in electron orbitals) can be assembled and a definite and nonrandom distribution becomes clear. In the case of the Dodgers, these functions fell quite frequently beyond the bounds of the playing surface. In the case of the Yankees, the result of swinging the bat was that all too often the probability function for position of the batted ball reached a maximum in the glove of the catcher.

KENNETH R. MILLER

Brown University,
Providence, Rhode Island, USA

Another contender

SIR — I note the absence of contributions on systems ecology to your debate on evolution.

The branching tree is a forceful image of evolution which has proved reasonably compatible with the limited evidence of the fossil record. Recently, some palaeontologists engaged on rigorous studies of fossil lineages have come to crave some theoretical innovation which would better model the history of evolutionary progress¹. They may be suspected of trying to induce innovation by creating what they consider to be favourable conditions for an ideological revolution: that is, by promoting a theory of "punctuated equilibrium" as an antithesis to "gradualism". So far, the rather unsurprising product gleaned from articles in your journal is that evolution can occur at varying rates^{1,2}. It also appears that evolutionary progress by

punctuation of an equilibrium does not require a qualitatively different type of evolution from that predicted by the genetic theory².

Rates of evolution seem to depend largely on the rate of change in selective stress, which results from a changing environment: for example, the gradual evolution of a species supposes gradual change in environment. However, the range of physical environments has remained fairly uniform through time: their geography may change but this is usually accommodated by a mobile biota. New forms which evolve in small isolated populations under environmental stress are likely to be short-lived: when their isolation breaks down and normal conditions return they are likely to be out-competed by more broadly adapted and numerous "ancestor" species.

In contrast, the biological environment has a capacity for accretionary change³. "Gradualists" should look to the progressive development of biological systems for their selective pressures. Proponents of punctuated equilibrium might investigate the instabilities which are likely to be inherent in such systems.

GILES MORRELL

Calgary, Alberta, Canada

1. Williamson, P.G. *Nature* 293, 437-443 (1981).
2. Cronin, J.E., Boaz, N.T., Stringer, C.B. & Rak, Y. *Nature*, 292, 113-122 (1981).
3. Sepkoski, J.J. Jr, Bambach R.K., Raup, D.M. & Valentine, J.W. *Nature* 293, 435-437 (1981).

Is this a record?

SIR — A recent paper by Sadler¹ shows that the stratigraphic record of geological time is very incomplete. In commenting on this, van Andel² notes that contemporary evolutionary theorists ascribe major evolutionary steps to a jump advance or "punctuation" — concentration of major change in a brief time interval.

If major change happens rapidly, and rapid events are rarely preserved in the fossil record, it follows that most major evolutionary events are beyond study because they have no stratigraphic record. But the actual evidence for a punctuational view of evolution depends on literal reading of the fossil record, that is, the assumption that gaps between successive stages of the stratigraphic record are insignificant³. If gaps represent more time than fossiliferous strata do, gradual transitions between species will necessarily appear punctuated.

Sadler himself noted: "In their treatment of the fundamental character of biological evolution, Eldredge and Gould (1972) are concerned with the completeness of the record. Unfortunately they tackle only the question of the abundance of fossils. It is equally important to determine whether the completeness of the sediment accumulation is adequate for the recognition of 'punctuations' in the fossil record that are phyletic in origin rather than indicators of sedimentary lacunae. We have attempted to quantify the sedimentary aspect; these numbers need to be matched by estimates of the duration implied by 'punctuations'."

PHILIP D. GINGERICH

Museum of Paleontology,
The University of Michigan, USA

1. Sadler, P.M. *J. Geol.* 89, 569 (1981).
2. van Andel, T.H. *Nature* 294, 397 (1981).
3. Eldredge, N. & Gould, S.J. in *Models in Paleobiology* (ed. Schopf, T.J.M.) (Freeman-Cooper, San Francisco, 1972).

NEWS AND VIEWS

Reservoirs and the triggering of earthquakes

from Peter J. Smith

IN spite of early doubts, it is now well established that large reservoirs can generate earthquakes. When the possibility of reservoir-induced seismicity was first mooted during the mid 1960s, it was greeted with scepticism in some quarters, largely because comparable before-and-after records for the few examples then known were unavailable. Systematic seismic surveys had only been started after impounding was under way or even, in some cases, complete; and so the details of the pre-reservoir seismic background were uncertain at best. A scientifically rigorous presentation of the evidence was therefore impossible, even though in one spectacular case (Koyna, India, 1967) a large reservoir had clearly given a relatively aseismic zone the ability to kill.

But there are no doubts now, for well documented examples of reservoir-induced seismicity are known throughout the inhabited continents. In general terms, what happens is that the huge mass of water in a reservoir changes the stress patterns in the underlying rocks. The water load exerts a simple vertical pressure on the rocks immediately, whilst the pore pressure in them increases gradually as the water infiltrates at a rate that depends on their permeability and prior water content. In addition, the water reduces the rocks' coefficient of initial friction and in the longer term can change their chemical, and hence physical, properties by weathering. Even so, it is widely held that such effects are not by themselves sufficient to generate earthquakes in zones not already subject to tectonic stress. The role of the water is rather to trigger seismic events in a highly stressed system close to break point.

This is clearly the case in the Tadzhikistan region of the USSR, where the Nurek reservoir has been generating seismic activity since filling began in the early 1970s. In this, as in a number of other cases, the induced seismicity is not related directly to the absolute amount of water present but to variations in it. Bursts of

seismic activity are triggered when the water level changes, the intensity of the activity increasing with the rate of level change. The temporal fluctuations in the seismicity are thus easily related to the fluctuations in the phenomenon inducing it.

The spatial distribution of the seismic activity, by contrast, is less straightforward. Most of the induced earthquakes occur beneath the central part of the reservoir and upstream (southward) from it. The region containing that half of the reservoir immediately behind the dam, on the other hand, is almost completely aseismic. Yet both the seismic and aseismic zones are lithologically similar, consisting largely of Upper Cretaceous to Palaeogene shallow-water limestones interbedded with mixed components of shale and gypsum. What, then, governs the location of the induced seismicity?

The answer, as Leith *et al.* have discovered (*Geology* 9, 440; 1981), lies in the varying structure and the different permeability regimes to which it leads. The limestone units in the region beneath the reservoir are highly fractured and thus relatively permeable, whereas the gypsum and shale units have very low permeability. The alternation of permeable and impermeable strata thus gives rise to a strongly directional overall permeability. Normal to the bedding the overall permeability is very low because the largely impermeable layers prevent, or severely restrict, water transport. Parallel to the bedding, however, the permeability is high because the water can easily diffuse along the permeable layers. In other words, in their originally horizontal configuration the strata would have prevented most vertical fluid migration.

The strata beneath the Nurek reservoir are no longer horizontal, although

immediately upstream from the dam the effect is little different. Here the alternating layers of limestone and gypsum-shale are folded into a syncline. Any water entering one of the permeable layers could migrate to the bottom of the basin formed by that layer but would be prevented from moving any further in the vertical direction by the impermeable rock layer beneath. In this region, therefore, water cannot enter the vast mass of rock below and cannot thereby change the pore pressure. The reservoir here acts only as a load, which is insufficient to induce seismicity.

Further upstream from the dam, on the other hand, the strata are folded into a moderately plunging open anticline with the dip in the upstream direction. In this region water entering a permeable layer can diffuse along it and thus be carried to various depths without restriction. Here the pore pressure at depths does change and is sufficient to generate seismicity.

What all this amounts to in the end, of course, is yet another version of the old, old story, namely, that general geological processes governed by universal physical laws will nevertheless lead to unique consequences depending on the particular environment in which the processes take place. As far as reservoirs are concerned, it is known that the large masses of water involved can generate earthquakes; but whether and precisely where they will actually do so depends critically on local conditions.

The Nurek case shows that the pattern of induced seismicity rests upon very local conditions indeed. The lithological environment in the Nurek area is not unusually inhomogeneous on a regional scale, yet the distribution of seismic activity varies on a subreservoir scale. The obvious conclusion is that if the environmental effects of future reservoir impounding are to be predicted accurately, it will be necessary to investigate the rocks underlying the sites in no less detail. □

Peter J. Smith is editor of 'Open Earth' and is a Reader in the Department of Earth Sciences, The Open University, Milton Keynes MK7 6AA, Buckinghamshire, UK.

How eggs breathe while avoiding desiccation and drowning

from Jared M. Diamond

AN avian egg is an almost self-contained life-support system providing all the nutrients, minerals and water necessary for the embryo's development. All that is required from the outside is oxygen and heat. The design of an egg involves numerous compromises. It must be strong enough to resist accidental breakage, yet fragile enough for the embryo eventually to break out. It must lose exactly the right amount of water during incubation: neither so much that the embryo dries out, nor so little that it drowns in its own metabolic water. The egg must be sufficiently impermeable to prevent drying out, yet sufficiently permeable that the embryo can acquire oxygen and get rid of carbon dioxide. The egg's incubation time must be geared to its metabolic rate, size and parents' time budget. The egg's size and nutrient content must be geared to the embryo's size and developmental state at birth, and perhaps to the clutch size. How is all this possible?

Answers to many of these questions have come from 13 years of work at the State University of New York at Buffalo, by Hermann Rahn, Charles Paganelli, Amos Ar* (of Tel-Aviv University) and their colleagues¹. Their interest in eggs began when a new student innocently asked how eggs breathe. As they had never thought about this problem, they began studies which eventually yielded the answer: by diffusion. The rate-limiting barrier to the egg's gas exchange is the egg shell itself, through which gas transport is diffusive, not convective. Oxygen, carbon dioxide and water cross the shell at rates proportional to their gas-phase diffusion coefficients. Since these gases follow identical diffusion paths, measurement of the egg's conductance to one (K_{O_2} , K_{CO_2} , or K_{H_2O}) lets one calculate the other two conductances. Water flux may be simply measured as the weight loss of an egg kept in a desiccator and egg conductances for 61 species of duck² and 84 other species of birds³ have been measured in this way.

Diffusion occurs through microscopic pores arising from imperfect packing of the calcium carbonate crystals that make up the egg shell. The pores can be visualized by putting an egg in water in a closed jar connected to a vacuum pump: gas bubbles emerging over the whole surface of the egg mark the pores. Underneath the shell lie two membranes of protein fibre mesh-work, in turn resting on a tissue sheet called the chorioallantois in whose capillaries the avian embryo's gas exchange occurs.

Bird eggs vary in size from the 300 mg of a small hummingbird to the 9 kg of the recently extinct elephant bird of

Madagascar, and in incubation period from 11 to 80 days. It is therefore astonishing⁴ that all eggs lose 15 ± 2.5 per cent of their initial weight during incubation! As Ar and Rahn point out, the fixed water loss ensures that the egg maintains constant relative water content throughout incubation, in the face of metabolic consumption of dry matter (mainly yolk fat) and accompanying production of metabolic water. Without water loss, the embryo would drown. Yet why is egg water loss so constant among species, when yolk content varies with species from 14 to 67 per cent? One might guess, although it is as yet untested, that the critical variable is constancy of the embryo's tissue osmolarity — for an embryo sealed within its egg, the only means of osmoregulation is water loss.

But how do different eggs ensure similar water losses? The explanation, from measurements of egg K_{H_2O} in dozens of species, proves to be that K_{H_2O} increases in direct proportion to egg weight (not to egg surface area) and in inverse proportion to incubation time. K_{H_2O} is determined by knA_p/L , where n is the number of pores, A_p the mean area of individual pores, L the pore length or shell thickness and k a constant. The shell and its pores are synthesized just before laying and do not change during incubation. Bigger eggs have thicker shells and lower pore density but wider pores and greater fractional pore area. A 10-fold increase in egg mass is associated with a 2.7-fold increase in shell thickness, an 18-fold increase in total pore area and thus a $18/2.7 = 6.5$ -fold increase in K_{H_2O} .

Eggs of the Wedge-tailed Shearwater (*Puffinus pacificus*) and chicken, for example, are of similar size (58 vs 54 g), pore radius (7.6 vs 8.0 μm) and shell thickness (0.27 vs 0.35 mm), and achieve similar water losses by the time of hatching (14 vs 15 per cent), despite the fact that the incubation period is two and a half times longer in the shearwater (52 days) than in the chicken (21 days). This result is achieved through the shearwater having many fewer pores (3,700 vs 12,400 per egg), hence lower water conductance⁵ (6.1 vs 14.4 mg H_2O per day, torr). Chickens and Red-winged Blackbirds living at high altitudes, where water loss would otherwise be high because of low atmospheric pressure and high gas diffusion coefficients, have shells with lower total pore area and lower K_{H_2O} than sea-level populations, and hence have similar water loss during incubation⁶.

Regulation of water loss during incubation implies regulation not only of shell K_{H_2O} but also of nest humidity. Continual release of water and CO_2 tends to humidify the nest as well as to raise nest

P_{CO_2} . Water vapour pressure in the nest air is always above ambient, as Rahn, Ackerman and Paganelli⁷ documented by an ingenious 'egg hygrometer'. This consisted of an egg emptied of its contents, refilled with desiccated silica gel, placed in the nest under the brooding adult, weighed after several days and calibrated by weighing after storage in desiccators at known relative humidities. The result: all nests were maintained at water vapour pressures of 18–26 torr, although the species studied ranged from Alaskan gulls and Mexican desert island gulls to ducks and songbirds. Does the brooding adult measure nest microclimate and then ventilate the nest to maintain the required rate of egg water loss, ventilating more rapidly as ambient humidity rises towards nest humidity? Is the water vapour gradient chronically abolished in species nesting in humid habitats such as moss forest of tropical mountains?

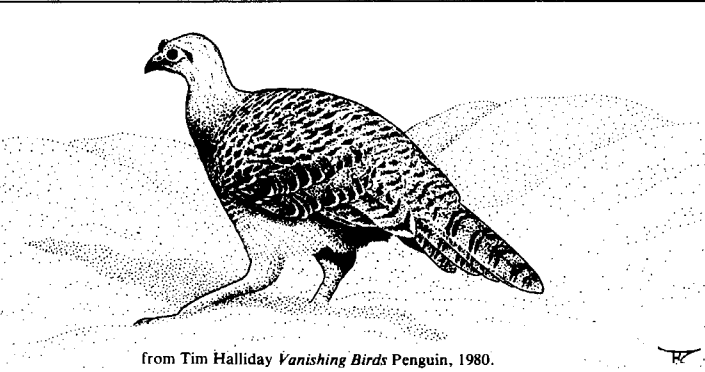
One group of birds violates these patterns: the mound-builders (family Megapodiidae) of New Guinea, Australia and adjacent islands. Unique among the world's birds, their eggs are not incubated by body heat but by external heat sources, including volcanoes and solar radiation. Several species bury their eggs in enormous mounds of rotting vegetation, which they build so that the heat of fermentation maintains a temperature of 37°C (see refs 8–10 and the figure). The humidity in these mounds is saturated. The water conductance of eggs of the mound-builder *Alectura lathami* is nearly triple that of a 'normal' egg of the same size and incubation period, due to an unusually thin egg shell¹¹. The megapode embryo avoids suffocating in the high CO_2 levels (9 per cent CO_2 !) and low O_2 levels (14 per cent O_2) in the mound, escapes drowning in its metabolic water, and still musters enough energy on hatching to spend two days digging its way up through a metre of dirt to the surface. The eggs of hole-nesting and burrow-nesting birds must also be able to deal with similar problems of gas exchange.

As one would expect, the metabolic rate of an egg (rate of O_2 consumption, Q_{O_2}) increases with egg size. But metabolic rate also decreases with incubation time; this is because water conductance decreases with incubation time, and as O_2 and H_2O traverse the shell by the same diffusion path, an egg of long incubation time must also have low O_2 conductance. Thus eggs of any size and incubation time not only lose a fixed relative amount of water (15 per cent) but also gain a fixed relative amount of O_2 , 100 cm³ per gramme of egg, over the

*In recognition of this work Rahn, Paganelli and Ar received the Elliot Coues Award at the 99th annual meeting of the American Ornithologists Union, Alberta, 24–28 August, 1981.

Jared M. Diamond is Professor of Physiology at the University of California Medical School, Los Angeles, California 90024.

The Mallee Fowl, one of Australia's mound-builders. In April or May, parent birds begin digging a pit that eventually measures 10 feet wide and 3 feet deep, and rake plant litter into it. With the early spring rains in August the litter begins to decompose and the birds cover it with a mound of sand up to 3 feet thick. The female lays eggs one at a time, in the mound, throughout the spring. The male is responsible for regulating the temperature of the eggs: during the spring he opens up the mound to let excess heat escape from the compost; in summer he piles more sand onto the mound during the day to insulate the eggs from excessive heat from the Sun, and each morning opens up the mound to dispel any heat remaining from the previous day; by autumn the compost has nearly stopped fermenting and the male opens up the mound each day so the eggs can be heated by the Sun, each evening he covers the eggs with warmed sand. By April all the eggs will have hatched.



from Tim Halliday *Vanishing Birds* Penguin, 1980.

incubation period^{12,13}. The metabolic rate is geared to the incubation time and delivers a finished chick by hatching time, whether 11 or 80 days after laying time.

The energy for this metabolism comes, of course, from the yolk. But not all eggs are equally endowed. Altricial eggs (ones whose chick hatches in a helpless state, at an early stage of development) contain only 24 per cent yolk, so that the egg is 84 per cent water, only 6 per cent lipid and offers only 1.1 kcal energy per g weight. Precocial eggs (whose chicks hatch able to fend for themselves, at an advanced stage of development) contain on the average 40 per cent yolk (67 per cent in a mound-builder), consist of 75 per cent water and 10 per cent lipid, and offer 1.9 kcal energy per g weight¹⁴. A parent can give its chick a head-start either by producing an egg with higher yolk content or just by producing a larger egg. Conversely, if a head-start is not important, a typical altricial egg saves its mother 80 per cent of the cost of manufacturing a typical precocial egg, through reduction in egg size and yolk content.

The questions posed by eggs seem endless. Large birds tend to lay large eggs, but egg size varies fivefold among birds of similar size belonging to different orders¹⁵. Why, among parents weighing 100 g, does a parasitic cuckoo produce a 4.5-g egg, an owl an 11-g egg and a petrel a 21-g egg? Evolutionary biologists debate this question, as well as the puzzle why the variation in incubation time from 11 to 80 days includes a large component independent of egg size, and the question of which variable among egg size, metabolic rate, incubation

time and developmental state at hatching ultimately determines the others. Bioengineers debate why the relation between total weight and skeletal weight has a similar slope but a different intercept among bird eggs, molluscs, spiders, land mammals and

whales¹⁶. Students of insects, reptiles, amphibia, fish and the duck-billed platypus have barely begun to explore for eggs of these animals the questions that Rahn, Paganelli, Ar and colleagues have illuminated for bird eggs. □

Sexual differentiation and H-Y antigen

from P.N. Goodfellow and P.W. Andrews

AMONG the many problems still awaiting the application of the new techniques of molecular genetics is that of mammalian sexual differentiation. The question can be simply phrased – why are males male and females female? Analysis of individuals with abnormal numbers of sex chromosomes indicates that the presence of the Y chromosome usually correlates with primary male characteristics¹. It has been suggested that this correlation is due to expression of the H-Y antigen controlled by the Y chromosome^{2,3}.

The study of the 'H-Y antigen' began with the observation that females from inbred strains of mice can reject skin transplanted from male mice of the same strain⁴. As the only qualitative genetic differences between the mice is that the Y chromosome is present in males but not in females, the antigen responsible for the graft rejection was termed histocompatibility-Y, or H-Y, antigen⁴. Although it is possible to use the transplantation assay to demonstrate that simple hormonal differences are unlikely to be responsible for the expression of H-Y antigens by males⁶, detailed investigation is hampered by difficulties in quantification and limitations in the use of transplantation assays in outbred species, such as man.

Serological definition of H-Y antigens

An important advance was made when Goldberg and colleagues showed that sera from female mice which had rejected male skin grafts are cytotoxic for sperm in the presence of complement, and that this activity can be absorbed by male but not female mouse cells⁷. The antigen(s) recognized serologically in this way is also called H-Y, although it has not been shown whether the same determinant is being

detected in the two different assays. Indeed, in a mutation experiment Melvold *et al.* have observed one male mouse, unfortunately sterile, that appeared H-Y negative by the transplantation assay, but was serologically H-Y positive⁸. Unfortunately the only cells that are susceptible to complement-mediated lysis by these anti-H-Y sera are sperm, which are difficult to read in the usual cytotoxicity assay, and tail skin epidermal cells⁹, which are difficult to obtain. Nevertheless many other cell types specifically absorb this activity, and by the absorption of syngeneic murine female anti-male sera followed by testing the residual activity on mouse sperm, serologically detected H-Y antigen is found to be strongly conserved and to be expressed on the male cells of all mammalian species tested. Conservation of this determinant is also seen in animals other than mammals where H-Y antigen is found to be associated with the heterogametic sex (in mammals, males are heterogametic having X and Y sex chromosomes, whereas in birds the female is heterogametic having the so-called ZW sex chromosome constitution)¹⁰.

This conservation and the association with one or other sex led to the suggestion that H-Y antigen was concerned in directing the organogenesis of the heterogametic gonad². Evidence for this has been presented and discussed elsewhere³. Suffice it to say here that in mammals the presence of testicular tissue has generally been shown to be associated

P.N. Goodfellow is in the Imperial Cancer Research Fund Laboratories, Lincoln's Inn Fields, London WC2A 3PX, and P.W. Andrews is in the Wistar Institute, Philadelphia, Pennsylvania 19104.

1. Rahn, Ar & Paganelli *Scient. Am.* **240**, 46 (1979); *Am. Zool.* **20**, 325 (1980).
2. Hoyt, Board, Rahn & Paganelli *Physiol. Zool.* **52**, 438 (1979).
3. Ar & Rahn in *Respiratory Function in Birds, Adult and Embryonic* (ed. Piiper, J.) (Springer, Berlin, 1978).
4. Ar & Rahn *Am. Zool.* **20**, 373 (1980).
5. Paganelli *et al. Physiologist* **23**, 152 (1980).
6. Rahn, Carey, Balmas, Bhatia & Paganelli *Proc. natn. Acad. Sci. U.S.A.* **74**, 3095 (1977).
7. Rahn, Ackerman & Paganelli *Physiol. Zool.* **50**, 269 (1977).
8. Frith *The Mallee-Fowl* (Angus & Robertson, Sydney, 1962).
9. Balin Z. *Tierpsychol.* **26**, 524 (1969).
10. Seymour & Ackerman *Am. Zool.* **20**, 437 (1980).
11. Seymour & Rahn in *Respiratory Functions in Birds, Adult and Embryonic* (ed. Piiper, J.) (Springer, Berlin, 1978).
12. Hoyt & Rahn *Respir. Physiol.* **39**, 255 (1980).
13. Rahn & Ar *Am. Zool.* **20**, 477 (1980).
14. Carey, Rahn & Parisi *Condor* **82**, 335 (1980).
15. Rahn, Paganelli & Ar *Auk* **92**, 750 (1975).
16. Anderson, Rahn & Prange *Q. Rev. Biol.* **54**, 1139 (1979).

with the expression of H-Y antigen, irrespective of genetic (that is, XX or XY chromosomal constitution) or phenotypic sex. The converse case – the expression of H-Y antigen being associated with the presence of testicular tissues – is not always true. Two explanations for individuals being H-Y antigen positive but without any testicular differentiation have been offered: either the H-Y molecule is 'functionally' absent (for example, the functional part of the molecule has been mutated while it remains immunologically active, or the H-Y receptor on the target cell has been inactivated) or testicular differentiation could depend on the quantity of H-Y antigens produced, a critical level required for the induction of testicular differentiation not being reached.

Technical difficulties with H-Y serology

The serological description of the H-Y antigen as outlined above is well established but this does not mean that H-Y serology is without difficulties. These result from two technical problems. The first is due to the sperm cytotoxicity test, which is one of the most difficult serological assays to perform. The extreme reactivity of sperm with naturally occurring heterophile antibodies requires careful control not only of the H-Y antisera but also screening and absorption of the serum used as a source of complement. Even the scoring of the sperm cytotoxicity test is difficult: sperm are mobile and very small and it is not easy to distinguish between dead and live sperm. In addition, there is often only a small difference between a positive and a negative result as the maximum kill in the assay is usually only 50–60 per cent, whilst the background kill in absence of any sera may be as high as 30 per cent. Unfortunately, the alternative assays used to detect H-Y antigen, which include every imaginable variation of cell surface immunological detection techniques¹¹, all seem to suffer from similar problems of high backgrounds and low maximum reactions. Undoubtedly this is due in part to the low surface density of H-Y antigen on many cells but it is also due to the poor-quality low titre of the anti-H-Y sera. Even after repeated immunization many female mice fail to make H-Y antibodies and those mice which do respond produce low titre antisera. This leads to the second problem – the small amounts of antisera that are available.

The paucity of H-Y antisera has two adverse effects. The first is a reduction in the exchange of antisera which is needed to standardize anti-H-Y reagents and to demonstrate that all H-Y assays are recognizing the same antigen(s). The second effect is to produce an understandable reluctance to perform quantitative absorptions (that is, multiple absorptions with varying numbers of cells), rather than the single point absorptions usually performed. Quantitative absorptions are essential for the proper quantification of the amounts of H-Y

antigen present. This is especially important when studying cells with postulated intermediate amounts of antigen, such as has been proposed to explain the recessive inheritance of male sex-determining genes in the polled goat and in humans^{12,13}.

An additional problem arises with the recently introduced H-Y test based on the complement-mediated killing of Raji cells (a male human, Burkitt's lymphoma line) by antisera from female rats immunized with male rat cells¹⁴. In this xenogeneic system absorption of anti-human heterophile antibodies is essential to ensure meaningful results. Furthermore, even if this absorption is carried out satisfactorily, there is no assurance that the antigen detected by the male-specific absorption of the remaining antibodies is the same as the antigen detected by absorption of syngeneic mouse serum tested on sperm, merely because they both appear to be male associated.

Conflicting results of H-Y typing

The technical difficulties and differences in methodology described above may explain the apparent contradictory results reported by the group in Freiburg associated with U. Wolf and U. Muller and by the workers in New York associated with S. Wachtel and G. Koo. The German group has demonstrated that females with Turner's syndrome (those lacking one of the two X chromosomes found in normal females) express H-Y antigen, although at reduced levels compared with males¹⁶. A similar situation has also been described in mice¹⁷. Based on these results Wolf and his colleagues have suggested that the H-Y antigen is coded by a structural gene on an

autosome, that this gene is induced by an activator coded by a Y-linked gene, and that it is repressed by the product of another regulator gene located on the short arm of the X chromosome which escapes X-inactivation in normal females. They further propose that interaction of these genes can lead to continuously variable levels of H-Y expression, depending on genotype, and that a certain threshold level of H-Y is required to induce testicular differentiation^{16,18}. One problem is that they would predict that an excessive number of X chromosomes in a Y chromosome-carrying individual would produce sufficient H-Y repressor to overcome the activator produced by the Y chromosome; such an individual should exhibit female characteristics. Nevertheless 49 XXXXY individuals with apparently male phenotype have been observed¹⁹.

Contrary to these ideas is the older hypothesis that the H-Y antigen is specified by a gene or genes normally located on the Y chromosome. H-Y antigen-positive XX males, such as are caused by the *Sxr* gene in mice, or the *Polled* gene in goats, are explained by a postulated translocation of H-Y genes from the Y chromosome to an autosome, a suggestion which has received support from molecular analysis of sex chromosome-specific DNA sequences in sex-reversed mice²⁰. Threshold levels for H-Y antigen activity and, perhaps multiple H-Y genes, are needed to explain why the sex-reversal properties of *Sxr* in mice are dominant whereas they are recessive in the case of the *Polled* gene in goats.

Consistent with Y-linkage of the H-Y structural gene is the finding that female Turner's syndrome patients with deleted or rearranged Y chromosomes can be either H-Y antigen positive or negative²¹. In particular, two patients have been described that have a Y isochromosome of the long arm. These patients are H-Y antigen negative^{21, 22}. It is difficult to understand why XO Turner's syndrome patients are H-Y antigen positive if XYp-Turner's syndrome patients are H-Y antigen negative.

Another possible point of contention between the two groups is the recent suggestion by the German group that XY fertile female wood lemmings do express H-Y antigen²³. In this species karyotypic analysis has demonstrated that some fertile females have a Y chromosome and an abnormal X chromosome, designated X*. As these individuals are normally fertile, wild populations of wood lemmings contain XX, XX* and X*Y females. The earlier studies of Wachtel *et al.* suggested that X*Y wood lemming females are H-Y antigen negative and that the abnormal X* chromosome carries a suppressor of H-Y antigen expression²⁴. The recent experiments yielding the opposite results (X*Y females being H-Y antigen positive) have led Wolf and Muller to propose that H-Y antigen affects its target cells via a receptor that is specified by an X-linked gene and that this gene has been deleted from the abnormal X

1. Davis, R. *J. med. Genet.* **18**, 161 (1981).
2. Wachtel, S.S., Ohno, S., Koo, G.C. & Boyse, E.A. *Nature* **257**, 235 (1975).
3. Ohno, S. *Monogr. Endocr.* **11**, 1 (1979).
4. Eichwald, E.J. & Silmsker, C.R. *Transplant Bull.* **2**, 148 (1955).
5. Billingham, R.E., Silvers, W.K. & Wilson, D.B. *Proc. R. Soc.* **163**, 61 (1965).
6. Silvers, W.K., Billingham, R.E. & Sanford, B.H. *Nature* **220**, 401 (1968).
7. Goldberg, E., Boyse, E.A., Bennett, D., Scheid, M. & Carswell, E.A. *Nature* **232**, 478 (1971).
8. Melvold, R.W., Kohn, H.L., Yerganian, G. & Fawcett, D.W. *Immunogenetics* **5**, 33 (1977).
9. Scheid, M., Boyse, E.A., Carswell, E.A. & Old, C. *J. exp. Med.* **135**, 938 (1972).
10. Wachtel, S.S., Koo, G.C. & Boyse, E.A. *Nature* **254**, 270 (1975).
11. Koo, G.C. *Hum. Genet.* **58**, 18 (1981).
12. Wachtel, S.S., Basurur, P. & Koo, G.C. *Cell* **15**, 279 (1978).
13. de la Chapelle, A., Koo, G.C. & Wachtel, S.S. *Cell* **15**, 837 (1978).
14. Fellous, M. *et al. J. exp. Med.* **148**, 58 (1978).
15. Wachtel, S.S. *et al. New Engl. med. J.* **293**, 1070 (1975).
16. Wolf, U. *et al. Hum. Genet.* **54**, 315 (1980).
17. Engel, W., Klemme, B. & Ebrecht, A. *Hum. Genet.* **57**, 68 (1981).
18. Wolf, U. *Hum. Genet.* **58**, 25 (1981).
19. Simpson, J.L. *Disorders of Sexual Differentiation and Clinical Delineation* (Academic, New York, 1976).
20. Jones, K.W. & Singh, L. *Hum. Genet.* **58**, 46 (1981).
21. Koo, G.C. *et al. Science* **198**, 940 (1977).
22. Rary, J.M., Cummings, D.K., Jones, H.W. & Rock, J.A. *J. Heredity* **70**, 78 (1979).
23. Wolf, U. & Muller, U. *Hum. Gene Mapping* **6** (in the press).
24. Wachtel, S.S. *et al. Nature* **264**, 638 (1976).
25. Koo, G.C., Chaganti, R. & Hammerling, U. *Hum. Genet.* **57**, 64 (1981).
26. Koo, G.C. & Varano, A. *Immunogenetics* **14**, 183 (1981).
27. Simpson, E. & Matsunaga, T. *Transplantation* **27**, 295 (1979).
28. Goodfellow, P.N. & Tippet, P. *Nature* **289**, 404 (1981).

chromosome of the wood lemmings²³.

Two explanations may be offered for these opposite results. Either there are two populations of wood lemming such that both H-Y⁺ and H-Y⁻ X*Y females occur, or, because the authors are using different antisera and different techniques, they are actually studying different antigens.

Future prospects

As in many other fields, monoclonal antibodies potentially offer a solution to the many problems of H-Y antigen serology. Monoclonal antibodies, which could be freely exchanged between different groups, will undoubtedly form the basis of agreement for the serological definition of the

H-Y antigen. One such monoclonal antibody has recently been described²⁵ and attempts have been made to demonstrate identity between the antigen recognized and H-Y defined by cell-mediated cytotoxicity²⁶, an *in vitro* assay which correlates well with transplantation rejection²⁷. Other monoclonal antibodies, which apparently recognize products controlled by the Y chromosome, are being studied in several laboratories (ref. 28; S. Wachtel, personal communication; M. Fellous, personal communication). Next year we should know much more about serologically defined antigens controlled by the Y chromosome. □

Electrostatic interactions in proteins

from Janet M. Thornton

PROTEINS are very polar molecules. Almost two-thirds of the twenty amino acid side chains are polar and four are charged at neutral pH. In addition, the arrangement of electrons on the backbone peptide groups gives rise to dipoles which, in the α -helix, align to give a helix macrodipole equivalent to a half unit of charge at each end of the helix (see ref. 1 for a review). In analysing protein structure, attention has been focused on the interactions of polar groups with water, a fundamental rule being that polar groups are exposed to solvent, or, where buried, form hydrogen bonds or salt bridges. By contrast, interactions between charged groups and dipoles within proteins have received little attention. It has been assumed that because all the ionized groups are exposed to water, the electrostatic contribution to free energy will be small, if not zero². Two papers^{3,4} recently published in *Nature* suggest, however, that the electrostatic interactions between charges and dipoles make important contributions to protein stability and may play a significant part in determining protein structure.

Wada and Nakamura³ consider the relative distribution of positive and negative charges (including the equivalent helix dipole charges) in fourteen proteins. For every possible pair of charges e_i , e_j separated by r_{ij} in the protein, they calculate the electrostatic interaction, $e_i e_j / r_{ij}$, and sum pairwise interactions within a given separation range. By comparing the distributions as a function of r for attractive and repulsive interactions they find that there are considerably more attractive, than repulsive, interactions at small separations. In other words, a charged group, including the apparent charges of the helix dipole, is on average

surrounded by charges of the opposite sign. This had already been shown to be true for myoglobin⁵ but, surprisingly, it is the first unequivocal proof that charged groups in general are not distributed randomly on the surface of proteins. The suggestion is that the charges, including the α -helix dipoles, contribute to the stability of the protein. This is an important result because it does not depend on dielectric constant, which is notoriously difficult to evaluate reliably for proteins.

The problem of how much energy the charge interactions contribute to the protein is much more difficult. Wada and Nakamura evaluate the electrostatic energy of the interactions using the method of Tanford and Kirkwood⁶ with the internal dielectric constant set to 4 and the external dielectric constant that of water, at 80. The electrostatic energies they obtain for the fourteen proteins range from +5 kcal mol⁻¹ for trypsin to -595 kcal mol⁻¹ for liver alcohol dehydrogenase. The model does not, however, adequately allow for the different dielectric environments felt by side chains in a complex protein structure. Many charged groups are completely exposed to solvent and their electrostatic interactions will be correspondingly decreased. Gurd and co-workers⁷ have tackled this problem by incorporating a solvent accessibility parameter which reduces electrostatic interactions between exposed groups. Their calculation for charge-charge interactions in myoglobin reduces the calculated electrostatic free energy from approximately -400 kcal mol⁻¹ to -8.5 kcal mol⁻¹. This energy contribution is more reasonable, although the lack of comparable experimental data is a problem.

In the hydrophobic interior of the protein electrostatic interactions are much stronger than those between charges exposed to solvent on the surface of the protein. Hol *et al.*⁴ use this argument to

suggest that the interactions between the helical dipoles, which are predominantly buried, make a significant contribution to protein stability. Their main observation is that a pair of anti-parallel helices is energetically more favourable than a pair of parallel helices, because of the electrostatic interaction between the peptide dipoles. The energy difference is evaluated by using the formal charges, assigned to the peptide backbone to mimic the observed dipole moment, and calculating the coulombic interaction $e_i e_j / r_{ij}$. They argue that a macroscopic dielectric constant is inappropriate at the level of atom-atom interactions in the hydrophobic core, and claim their energies are incorrect by no more than a factor of 2. Using this method, the anti-parallel arrangement is more stable, by almost 20 kcal mol⁻¹, than its parallel equivalent. They suggest that this electrostatic contribution may have a significant role in determining the observed preference for the anti-parallel arrangement of helices in the all-helical proteins. For four of these proteins the dipolar electrostatic stabilizations range from -5 to -23 kcal mol⁻¹.

In the α/β proteins with alternating α -helices and β -strands, the α -helices lie approximately parallel, maximizing unfavourable dipolar interactions. However, Hol *et al.* point out that in parallel β -sheets there is also some alignment of the peptide dipole, which gives a small macrodipole, equivalent to one-fifteenth of a unit of charge at each end of the strand. This dipole, together with the fact that the helices are usually tilted because of the twist of the sheet, effectively cancels out the unfavourable interactions.

Clearly the helix dipole interactions do contribute to the stability of the anti-parallel helices. However, the number of kilocalories involved has yet to be satisfactorily evaluated. The figures calculated by Hol *et al.* provide the maximum possible stabilization to be derived from dipole interactions. Most helices are partially, if not predominantly, exposed, especially in the small all-helical proteins, so presumably some of the peptides will be in contact with solvent. However, the closest contacts will be in the hydrophobic core, so although this effect should be considered, perhaps in terms of solvent accessibility, it may not markedly affect the energies.

The importance of the electrostatic dipole interactions for determining the anti-parallel alignment of helices and thus protein folding is difficult to assess. Many other factors are involved, whose energies may be even more difficult to quantify. Topological considerations are one such possibility. For every parallel interaction between two sections of chain there must be an anti-parallel connection. For example, in the parallel β -barrels, such as in triose phosphate isomerase, the helical connections run approximately anti-parallel to the strands. Therefore, we would expect anti-parallel connections to be more common. The dipole interactions

Janet M. Thornton holds an SERC Advanced Fellowship at Birkbeck College, Malet Street, London WC1E 7HX.

will help stabilize the anti-parallel arrangements, although since the dipoles must be buried to make any significant contribution to energy, they only become important once the hydrophobic core of the protein has been formed.

To conclude, both these papers draw our attention to the contribution which charged groups, including the α -helical dipole, make to protein stability. The difficulty of assigning an effective dielectric constant, or indeed if a macroscopic dielectric constant is valid for these atom-atom interactions, has yet to be satisfactorily solved and needs urgent attention. More experimental data for electrostatic energies or effective dielectric constant, such as that provided by Rees⁸, would be of enormous value. With high-resolution coordinates now available for many proteins, it should be possible to characterize charge interactions in detail.

A more exciting possibility is that charged groups, being on the surface of proteins, will be of general importance for recognition between molecules, including protein-DNA interactions. Preliminary evidence from the cytochrome *c* peroxidase structure⁹ and the assembly of tobacco mosaic virus subunits¹⁰ lends support to this hypothesis. Perhaps it is in this context that direct electrostatic interactions may prove most significant. □

1. Wada, A. *Adv. Biophys.* 9, 1 (1976).
2. Finney, J.L., Gellatly, B.J., Golton, I.C. & Goodfellow, J. *Biophys. J.* 32, 17 (1980).
3. Wada, A. & Nakamura, H. *Nature* 293, 757 (1981).
4. Hol, W.G.J., Halic, L.M. & Sander, C. *Nature* 294, 532 (1981).
5. Friend, S.H. & Gurd, F.R.N. *Biochemistry* 18, 4620 (1979).
6. Tanford, C. & Kirkwood, J.G. *J. Am. chem. Soc.* 79, 5333 (1957).
7. Friend, S.H. & Gurd, F.R.N. *Biochemistry* 18, 4612 (1979).
8. Rees, D.C. *J. molec. Biol.* 141, 323 (1980).
9. Poulos, T.L. & Kraut, J. *J. biol. Chem.* 255, 10322 (1980).
10. Bloomer, A.C., Champness, J.N., Bricogne, G., Staden, R. & Klug, A. *Nature* 276, 326 (1978).

Cell mediated immunity in theileriosis

from F.E.G. Cox

THERE are strong indications that antibody does not play a major role in the immune responses to parasitic protozoa that inhibit nucleated cells but there is some evidence that antibody-independent cell-mediated mechanisms may play a part in protection and recovery. Such mechanisms may involve lymphocyte mediated cytotoxicity (Class I reaction), soluble lymphocyte products (Class II reaction) or natural killer (NK) cells but these mechanisms, although well characterised in mice, are difficult to implicate in the parasitic diseases of man and domesticated animals. It is, however, now clear that cell mediated cytotoxicity plays a major, if not the major, part in the immune response to one of the most important protozoan diseases of cattle, a form of theileriosis known as East Coast Fever.

Theileriosis is caused by several species of *Theileria*, of which *T. parva*, in Africa, and *T. annulata*, which occurs in North Africa, the Mediterranean area and elsewhere, are the most important. Considerable mortality results, particularly in calves. Theileriosis is a tick-borne disease and cattle become infected when sporozoites are injected by an infected tick. The sporozoites enter lymphocytes in which they grow and undergo nuclear division giving rise first to macroschizonts and later microschorizonts, the products of which infect red blood cells and thence ticks once again. While in the lymphocytes the parasites cause the host cells to undergo rapid division transforming them into what are effectively lymphoblastic cell lines. During the stage of macroschizont production there is an accompanying massive destruction of lymphocytes and the infected animal becomes ill and may die.

Theileria sporozoites can also transform lymphocytes *in vitro*^{1,2} and the availability of such lymphoid cell lines has provided scientists with an ideal opportunity to study cell mediated immune responses against cells of the same genetic type, for uninfected cells can be taken from an animal before infection, transformed and maintained for subsequent study during the course of an infection³. Scientists at ILRAD (International Laboratory for Research on Animal Disease, Nairobi) have followed *T. parva* infections in groups of cattle that died from theileriosis, recovered naturally from the disease or were protected by being infected and treated at the same time with a long acting tetracycline⁴⁻⁷. In unprotected cattle, by day 14 after infection there was evidence of nonspecific cytotoxic activity against both infected cells and unrelated cells of a mouse tumour line⁵ suggesting a mechanism similar to the natural killer (NK) cell phenomenon seen in mice and humans⁸. In immunised cattle, on the other hand, small numbers of macroschizonts were detected during days 15-18 after infection and cytotoxicity was confined to genetically identical infected cells^{5,6}. An analysis of the mechanisms involved implicated thymus derived lymphocytes as the effector cells⁷ and a restriction mechanism whereby effector cells have to recognise both foreign antigens and self antigens simultaneously⁶. This is characteristic of class I cell mediated immune responses such as seen in virally altered cells⁸. Although the details of class I reactions and the nature of restriction in cattle remain to be

F.E.G. Cox is Professor of Zoology at King's College London.

100 YEARS AGO

A SPECIAL despatch has been received at St. Petersburg from M. Sullowsky, dated Irkutsk, December 26 (O.S.) 1881, which says: "At 10 o'clock on the morning of August 9 I parted with the *William Rodgers*, which shaped her course for Herald Island. The clipper *Strelok* then returned to the Chinese ports. Up to that time the *Strelok* and *William Rodgers* had kept company. They were joined in Providence Bay by an American schooner, having on board the captain of a whaler which had stranded. This captain narrated that he had seen a boat with dead men on board which had been driven upon Herald Island. The boat also contained, besides other articles, some silver spoons with the name *Jeannette* engraved on them. In consequence of this narrative the captain of the *William Rodgers* resolved to proceed to Herald Island with the view of wintering there, and, with the aid of dogs purchased in Kamschatka, sending out his crew in small parties to the various sides of the island and its vicinity to search for the lost explorers."

Admiral Mouchez will give his usual annual *soirée* at the Paris Observatory in March. He had distributed to the leading Parisian engineering firms the conditions for the construction of the cupola for the great equatorial to be built in the newly-annexed grounds. The diameter of the revolving cupola is to be 20 metres. The form must be hemispherical. The time required for rapid revolution is 10 minutes. It is to revolve in the same direction as the heavens, and the mechanism will cause the revolution of a seat for two astronomers. From *Nature* 25, 249 & 250; January 12, 1882.

determined all the evidence suggests that in theileriosis the main component of the immune response is the recognition of theilerial antigens on infected cells and the subsequent destruction of these cells by thymus derived lymphocytes.

The significance of this work is that it demonstrates the importance of cell mediated immunity in the natural host of an important protozoal disease and also extends the phenomenon of cytolytic restriction to the protozoa. Thus theileriosis joins leishmaniasis⁹ as examples of intracellular protozoa that are controlled by cell-mediated immune responses, class I in the case of the former and class II in the latter, giving support to the suggestion that the main function of the major histocompatibility complex (MHC) molecules is to guide T lymphocytes to deliver appropriate responses to diverse kinds of infection¹⁰. The diseases of man caused by *Toxoplasma gondii* and *Trypanosoma cruzi*, which are also intracellular, may soon yield up their secrets. □

1. Brown, C.G.D. *et al. Nature* 245, 101 (1973).
2. Pearson, T.W. *et al. Nature* 281, 678 (1979).
3. Emery, D.L. & Morrison, W.I. *Immunology* 40, 229 (1980).
4. Allison, A.C. & Eugui, E.M. *Amer. J. Path.* 102, 114 (1981).
5. Emery, D.L. *et al. Immunology* 43, 323 (1981).
6. Eugui, E.M. & Emery, D.L. *Nature* 290, 251 (1981).
7. Emery, D.L., Tenywa, T. & Jack, R.H. *Infect. Immunity* 32, 1301 (1981).
8. Doherty, P.C. & Zinkernagel, R.M. *J. exp. Med.* 141, 502 (1975).
9. Cox, F.E.G. *Nature* 291, 111 (1981).
10. Mitchison, N.A. *Immunology Today* 1, 91 (1980).

Oddities in the initiation of translation

from Peter Model

A report from Kastelein, Remaut, Fiers and van Duin in this issue of *Nature* (p.35) suggests that we may need to modify some of our ideas on how translation is initiated — at least in the particular case of the lysis gene of bacteriophage MS2. Most researchers have concluded that translation is initiated independently at each gene, even in multi-gene operons. It may be that mRNA structure will occlude the initiation site at a particular place, and in such instances translation of a preceding gene may be required for effective initiation at some cistrons, but this has generally been viewed as a secondary, complicating phenomenon which obscures the otherwise fundamentally independent nature of the initiation process.

Kastelein and his colleagues now present very persuasive evidence that initiation at the lysis gene requires translation of the preceding coat gene and a shift in the reading frame of at least some of the ribosomes into the +1 frame which, in turn, leads them to terminate at an *ochre* codon three nucleotides before the initiating AUG start codon. This termination is required for translation of the lysis gene for if the *ochre* codon is removed, translation is abolished. If by suitable manipulation of the cloned gene a frame shift is introduced which forces all or most of the ribosomes into the +1 frame, initiation is enhanced. Translation of the lysis gene is strictly dependent on active translation at the coat cistron; supplying coat protein *in trans* does not replace this requirement.

These data strongly suggest that there is something special about ribosomes that have just terminated at the *ochre* codon . . . that they are in a state which facilitates initiation at sites which would not normally be translational initiators. It is not yet clear what this special property might be; it could be a special state of the ribosome, or retention of the 30S ribosomal subunit on the RNA after termination of synthesis. It has been noted previously that termination at various chain-terminating codons within genes facilitates re-initiation at specific distal sites, although here too the mechanism(s) remains undefined.

These observations would at first glance seem to run counter to recent *in vitro* results from Kaji's laboratory. Ryoji, Berland and Kaji (*Proc. natn. Acad. Sci. U.S.A.* 78, 5973; 1981) report that when they removed the ribosome release factor from *Escherichia coli* ribosomes and then supplied an RNA containing an amber mutation as template, these release factor-depleted ribosomes synthesize two fragments of the relevant protein. One starts at the usual initiating methionine and stops at the *amber* codon, while the other starts at the codon immediately after the

amber codon and ends at the normal C terminus of the protein. Ryoji *et al.* argue that this result excludes 'special conformation' or 'sterile pass' models of ribosome initiation following termination. They conclude that when a ribosome terminates it either dissociates completely or else starts again at the very next codon. The characterization of the products of the release factor-depleted ribosomes was careful and thorough; the question that

Peter Model is in the Rockefeller University, 1230 York Avenue, New York, NY 10021.

remains is whether there is any relationship between the behaviour of these release factor-depleted ribosomes *in vitro* and the phenomenon described *in vivo* by Kastelein *et al.* Kaji and co-workers have carefully characterized an experimental system of their own design which may shed light on the mechanism of termination/initiation, but which may not be relevant to the genesis of the phenomena described by Kastelein *et al.* They propose on the basis of their rather specific *in vitro* work that there are only two possible fates for a ribosome when it reaches a stop codon: release with no special state, or retention on the mRNA, in which case it will simply start protein synthesis at the next codon. The *in vivo* studies from Van Duin's lab suggest that other possibilities are available, and have been exploited by the RNA phages. □

Magnetic messages from the Earth's core

from David Gubbins

OVER the centuries the Earth's magnetic field has shown marked changes. This 'secular' variation, known even to the early navigators, almost certainly originates in the core of the Earth and provides one of the few sources of information about the physical processes that occur there. Shorter-duration magnetic field changes can also be recorded (on a time scale of a few years) but are hard to interpret because of the difficulty of separating changes in intrinsic magnetism from the short-term influences of the Sun and the upper atmosphere. The major problem at present is how best to make use of the information provided by the secular changes.

Allredge¹ has recently proposed a model of intrinsic field fluctuation. The fluctuation propagates up through the Earth's mantle, which is a rather poor conductor of electricity, and his model produces surprisingly sharp changes at the surface. In consequence, if we are to decipher these signals from the interior of the Earth, we must measure and study even shorter-period wavelength variations than hitherto. This important conclusion stands quite independent of Allredge's specialized and rather arbitrary models, and it is timely coming soon after the MAGSAT mission which has provided the most complete and accurate magnetic chart we have ever had². Unfortunately MAGSAT only flew for a few months, not long enough to chart the secular changes with any accuracy. At present the best data for this purpose come from permanent magnetic observatories. A second MAGSAT, flown about five years hence,

would yield vastly more accurate and complete information.

The liquid core of the Earth extends out to about one-half the radius, is made mainly of iron, has a viscosity rather like water and is a better electrical conductor than copper. The mobility of this liquid causes the magnetic field to change through electromagnetic induction. The phenomenon occurs very rapidly in geological terms, the fluid moving at a rate of a few millimetres per second.

Secular change was much discussed, even in the earliest days of geophysics. Halley, like others of his time, believed the magnetic field to be due to magnetized rocks. The secular changes could not be due to motion of rocks near the surface because the required speed would be too great, so Halley proposed that the interior of the Earth was hollow, and that an inner, magnetized sphere was rotating westwards, generating the slow westward drift of the magnetic field pattern³. This remarkable theory came 200 years before the discovery of a liquid core by seismology. With modern ideas of a 'hot' Earth and a liquid iron core we can now deduce a value of the core radius from magnetic measurements alone⁴.

The big snag is that we do not yet fully understand the dynamo theory of the origin of the magnetic field itself, and therefore there are no really satisfactory models of the field changes in the core. Allredge uses a very common mathematical device: representation by a distribution of dipole sources. These dipoles have no physical meaning whatever, unless one wants to adhere to Halley's ideas of moving magnetized rocks and represent them by little bar magnets in the core, which is not the intention. Granted a source of field,

David Gubbins is in the Department of Earth Sciences, University of Cambridge, Cambridge CB3 0EZ.

there is still a serious theoretical difficulty in propagating magnetic disturbances through a liquid conductor, and in order to make progress Alldredge has been forced into treating the core as a solid, thus substituting the theory of electromagnetic propagation for the much more difficult hydromagnetic theory. But the physics of the two situations are very different: a liquid, for example, allows propagation of a variety of waves which are impossible in a solid, and these waves will allow magnetic variations to propagate, when in a solid they would be screened due to electrical conduction. In fact the core possesses a whole spectrum of free hydromagnetic oscillations⁵.

Although hydromagnetic theory is difficult, it is possible to describe the induction process in the core by Maxwell's laws. On the time scale over which we have observations, say 10–20 years, the electrical conductivity of the core is so high that

resistance effects can be ignored. This assertion can, and has been, tested against observation. One might hope, therefore, to infer the motion of the fluid at the top of the core from magnetic observations, but this is not possible⁶ unless we can confidently make further assumptions about the forces driving the flow. However, recent work has shown how to find some general features of the core flow even though a detailed map of the motions may be beyond our grasp. For example, upwelling of liquid from deep inside the core can be measured at points where the magnetic field at the core attains a maximum of minimum vertical component. Surprisingly, the upwelling is very small at these locations⁷, suggesting that the liquid is not convecting strongly, at least near its upper surface.

This and other recent proposals⁸ suffer from the poor quality of available data. Field changes are monitored at only a few

Southern Hemisphere sites. The field must be projected down onto the core surface before it can be interpreted — a procedure well understood in principle but inherently unstable when the observations are inaccurate or sparse. Two recent models of the secular change for 1965 agree quite well at the surface of the Earth but differ by factors of up to two or three at the core. The differences are due in part to the methods that are used in computing the models from the observations, and not all to inadequate data, so that while we wait for another MAGSAT there is still hope for improvements in models of past years by new analytical methods. □

1. Alldredge, L.R. *J. geophys. Res.* **86**, 7957 (1981).
2. Barraclough, D.R. *Nature* **294**, 14 (1981).
3. Halley, E. *Phil. Trans. R. Soc.* **17**, 470–478 (1692).
4. Hide, R. & Malin, S.R.C. *Proc. R. Soc. A* **374**, 15 (1981).
5. Hide, R. *Phil. Trans. R. Soc. A* **259**, 615 (1966).
6. Backus, G.E. *Phil. Trans. R. Soc. A* **263**, 239 (1968).
7. Whaler, K.A. *Nature* **287**, 528 (1980).
8. Barton, E.R. *Phys. Earth planet. Int.* **20**, 111 (1979).

Animal disease and conservation

from Geoffrey R. Smith

WHEN the material interests of man and the well-being of wild animals are considered side by side, real conflicts of interest emerge. Advances in human medicine lead to the growth of population, an increased demand for land and a corresponding loss in wildlife habitat. Environmental pollution is a well recognized side effect of industrial development. Less well understood by many is the threat to wild animals from infectious disease. Fatal or incapacitating infections may be introduced by domestic animals, and there can also be times when the destruction of wildlife becomes necessary to protect man and domestic animals from disease. These matters were considered at a recent symposium* held at the Zoological Society, under the general chairmanship of Sir Andrew Huxley, PRS.

From 1889 to 1898 rinderpest swept from North-east Africa to the west coast and Cape of Good Hope in a 'virgin-ground' panzootic that produced catastrophic losses in cattle and many wild ungulate species (W. Plowright, ARC Institute for Research on Animal Diseases). The disease was eliminated from southern Africa, but a focus of mild permanent infection later became established in the Serengeti region. Rinderpest disappeared from both cattle and game in that region in about 1964, after cattle had been treated with attenuated tissue-culture vaccine for several years. Between 1961 and 1971 the number of buffaloes and wildebeest in the Serengeti more than doubled, probably largely because the

young no longer died from rinderpest. There are now 1–15 million highly susceptible wild ungulates in the Serengeti. Once again they are under threat of invasion by rinderpest from the north-east. Plans need to be made now to meet this threat.

Although antigenically very similar, rabies virus strains from vampire bats in South America and foxes in western Europe have characteristic features detectable by monoclonal antibodies. Bats and foxes are sometimes killed as a means of controlling rabies. When fox population density falls below 0.3 per km² the disease disappears. An alternative strategy, still experimental but possibly of great significance, is the use of attenuated virus, concealed in chicken-head baits, to immunize foxes. The oral-vaccine virus multiplies in the tonsil and so far shows no sign of reversion to virulence (F. Steck, Vet.-Bakteriologisches Institut, University of Bern).

The role of animal infections in the origin and spread of human influenza pandemics is still the subject of much speculation (M. Kaplan, Pugwash Conferences on Science and World Affairs, Geneva). Inter-species infections can and do occur on occasion, apparently when mutations and recombinations take place. The segmental mode of replication of influenza virus favours recombinative encounters between human and animal strains. Wild birds, in which infection is mainly intestinal and symptomless, constitute an ineradicable reservoir of the everchanging virus — a reservoir that is a permanent potential threat to man and other animals. Recent observations on whales, seals and terns show that

influenza in wildlife is not always sub-clinical, and that high mortality can occur.

Botulism kills millions of waterfowl annually on lakes and marshes throughout the world, and no radical control methods are known (G.R. Smith, Institute of Zoology, The Zoological Society of London). Mud from certain localities, such as the Camargue, inhibits the growth of *Clostridium botulinum*. Our knowledge of the inhibitory mechanisms is incomplete, and a deeper understanding might give some hope for control. As an adjunct to wetland management, surveys for the causative organism are advocated to distinguish between areas of high and low risk.

The evidence that bovine tuberculosis infection is transmitted from badgers to cattle is irrefutable, and the impossibility of separating the two species physically or protecting them adequately by vaccination has necessitated the policy of gassing badgers in the south-west of England (Sir William M. Henderson). While research on the problem continues, the present control policy cannot be abandoned without causing suffering for us all — human beings, cattle and, not least, badgers.

It is impossible to generalize about the part played by disease in the control of numbers of wild animals in comparison with that due to water and food supply. But what the symposium made absolutely clear is that the whole field of animal disease is now of major importance. The subject overlaps at almost every point with that of disease in man and in domesticated animals and offers a great intellectual challenge to the research worker. □

Geoffrey R. Smith is Head of the Department of Infectious Diseases, The Institute of Zoology, Regent's Park, London NW1 4RY.

*The proceedings of the symposium on 'Animal disease in relation to animal conservation', held at the Zoological Society of London on 26 and 27 November 1981, will be published for the Society by Academic Press in the series 'Symposia of The Zoological Society of London'.

ARTICLES

Ion-supported tori and the origin of radio jets

M. J. Rees*, M. C. Begelman*[‡], R. D. Blandford[†] & E. S. Phinney*

* Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

[†] Theoretical Astrophysics, California Institute of Technology, Pasadena, California 91125, USA

[‡] Department of Astronomy, University of California, Berkeley, California 94720, USA

While apparently supplying tremendous power to their extended radio-emitting regions, the nuclei of most radio galaxies emit little detectable radiation. It is proposed that at the centre of each is a spinning black hole surrounded by a torus of gas too hot and tenuous to radiate efficiently. The torus anchors magnetic fields which extract rotational energy from the hole in the form of two collimated beams of relativistic particles and fields. These in turn drive the observed radio jets and hot spots. A large supply of accreting gas is thus unnecessary and radio galaxies may be interpreted as starved quasars.

ALL extended radio sources seem to be associated with elliptical galaxies or quasars, and the evidence¹ suggests that quasars too are located in ellipticals. As the present number-density of dead quasars is comparable with that of elliptical galaxies^{2,3}, it is reasonable to suppose that radio sources are powered by $\sim 10^8 M_\odot$ black holes³ left over from a quasar phase. A long-standing problem in the theory of active galactic nuclei has been that of finding a supply of nuclear gas adequate to 'feed the monster'. Stars in the galactic nucleus cannot provide more than $\sim 10^{-3} M_\odot \text{ yr}^{-1}$, unless conditions are so extreme that stellar collisions dominate the evolution⁴. Gas originating further out must dispose of substantial angular momentum to enter the nucleus^{5,6}, and it must do so before succumbing to galactic winds or star formation.

The observational evidence is less ambiguous: radio sources almost certainly do have low fuelling rates. Low-luminosity double-jet radio sources of the sort being mapped with the VLA generally have no detectable nuclear emission in the optical or X-ray bands; of the more powerful double hot spot radio sources, only a few (such as 3C390.3 with $\sim 2 \times 10^{44} \text{ erg s}^{-1}$ in both optical and X-ray bands⁷, and Cygnus A with $\sim 2 \times 10^{43} \text{ erg s}^{-1}$ in optical continuum⁸) have nuclear luminosities above $10^{42} \text{ erg s}^{-1} = 10^{-4} M_8^{-1} L_{\text{Edd}}$ (refs 9, 10). They cannot be produced by any process requiring super-Eddington luminosity, unless the hole mass is very small. This it cannot be because (even aside from the progenitor argument) some low-luminosity radio sources have giant ($\sim \text{Mpc}$) extensions with minimum energies¹¹ of up to $2 \times 10^{60} \text{ erg s}^{-1} = 10^6 M_\odot c^2$. Even if we make optimistic estimates of the efficiency of the central 'engine' and of the likely efficiency with which power emitted from the nucleus can accelerate relativistic electrons in the radio lobes, we must conclude that a mass of at least 10^7 – $10^8 M_\odot$ is involved (in computing the powers and energies above, we have used a Hubble constant of $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$).

Recent VLBI observations^{12–17} have corroborated the view that the jets which power extragalactic sources are collimated on scales $\leq 1 \text{ pc}$. These observations reveal small-scale structure which is generally aligned with larger-scale structure. Misalignments of up to $\sim 45^\circ$, which have been found in the core-dominated superluminal sources, are probably caused by viewing a slightly curved jet along an almost parallel line of sight. In the case of the cores of extended radio sources, however, the alignment is fairly good. Any mechanism which purports to explain them must, therefore, be capable of producing collimated jets of at least trans-relativistic velocities while radiating as waste heat much less than an Eddington luminosity. Furthermore, it must, at least in some more extreme sources, put orders of magnitude more power into the jet than escapes as (detectable) radiation from the nucleus. These constraints rule

out the class of models involving thick disks supported by radiation pressure^{18–25}.

Sub-critical accretion

To understand radio sources, we should therefore study accretion at very low rates, $\dot{M} \ll \dot{M}_{\text{Edd}} \equiv L_{\text{Edd}}/c^2$. Accreted gas should carry with it some magnetic flux, and shearing motions will maintain the magnetic field energy $B^2/8\pi$ at a significant fraction β^{-1} of the pressure P . The Larmor radius is then, for the applications which interest us, about 10 orders of magnitude smaller than the scale of the flow. Ordinary 'molecular' viscosity will consequently be suppressed, and magnetic (or possibly turbulent) viscosity will probably govern the shear stress. The viscous stress may be written as αP , where estimates of magnetic viscosity suggest $\alpha \geq 0.01$ (ref. 26). Although this estimate is uncertain, there is no reason why α should diminish as $\dot{M}$ falls: if α is fixed, the torque per unit mass will be independent of $\dot{M}$ (except insofar as the disk structure depends on $\dot{M}$). Thus the inflow time in a disk of half-thickness $h(r)$ around a hole of gravitational radius $r_g = GM/c^2$,

$$t_i = \alpha^{-1} \left(\frac{r}{r_g} \right)^{3/2} \left(\frac{h}{r} \right)^{-2} \left(\frac{r_g}{c} \right) \quad (1)$$

does not depend explicitly on density. (See refs 27–29 for the standard theory of thin disks with $h \ll r$.) The cooling time, on the other hand, is inversely proportional to the density.

The thickness of the disk depends on how fast it can radiate (or otherwise dispose of, say by a wind) the binding energy liberated by viscous stresses. The major cooling mechanisms available to the accreting matter are illustrated in Fig. 1. Depending on initial conditions and external perturbations the matter may find itself in any of several temperature states. In fact, for $10^{-7} \leq (\dot{M}/\dot{M}_{\text{Edd}})\alpha^{-2} \leq 50$ there are two self-consistent solutions to the fluid and radiation equations.

One is a thin, dense disk that remains at $\sim 10^4 \text{ K}$. If matter with well defined angular momentum is fed in steadily (with high angular momentum) at a large radius, and there are no traumatic perturbations, this cold disk should be selected. These disks are, however, subject to instabilities^{5,30,31}, and if heated up and thickened, they will thereafter be unable to radiate the energy dissipated by viscous stresses, and will inflate to become pressure-supported tori with vortices along the rotation axis (see Fig. 2). The hot torus solution will also be selected if cool matter is preheated by high-frequency non-thermal radiation (for example, from a jet), or if the gas is supplied hot (for example by colliding stellar winds⁶ or by tidal disruption of stars^{32,33}). Both of these latter examples lead to flows which join naturally onto the ones described here.

For an optically thin disk to radiate by bremsstrahlung, the energy dissipated between r and $r/2$, the accretion rate must satisfy

$$\dot{m} = \dot{M}/\dot{M}_{\text{Edd}} \geq \alpha_t^{-1} \left(\frac{m_p}{m_e} \frac{r_g}{r} \right)^{1/2} \alpha^2 \left(\frac{h}{r} \right)^2, \quad r > \left(\frac{h}{r} \right)^2 \left(\frac{m_p}{m_e} \right) r_g \quad (2a)$$

$$\dot{m} \geq \alpha_t^{-1} \alpha^2 \left(\frac{h}{r} \right)^3 (\log \gamma_{\text{re}})^{-1}, \quad r < \left(\frac{h}{r} \right)^2 \left(\frac{m_p}{m_e} \right) r_g \quad (2b)$$

where $\alpha_t = 1/137$ is the fine structure constant and $\gamma_{\text{re}} = 3 k T_e / m_e c^2$. Hence at the low accretion rates of interest ($\dot{m} \approx 10^{-4}$), a disk once heated to $T \approx T_{\text{virial}} = (m_p c^2 / 3k)(r_g/r)$ will be unable to cool through bremsstrahlung and must remain pressure-supported with $h \approx r$.

As the gas moves in, however, it will shear and amplify any seed magnetic field until $B^2/8\pi = P_{\text{gas}}$, and the fields begins to rise buoyantly out of the disk (moving perpendicular to the plane of a thin disk; towards the vortex funnel in a thick torus). We thus expect the ratio β to be self-limiting at $\beta \approx 1$, although β may be $\ll 1$ in a corona if the latter is threaded by field lines anchored to interior (high-pressure) parts of the disk, as in the solar atmosphere. The corona may also join onto a wind (M.C.B. *et al.* in preparation, and ref. 34). This magnetic field provides the torus with a new cooling mechanism: for $r < (m_p/m_e)r_g$, electrons at the virial temperature of the ions would be relativistic, and each would radiate an energy

$$\Delta E_t \approx k T_i \left(\frac{50 r_g}{r} \right)^{31/4} \dot{m}^{-1/2} \alpha^{-3/2} \beta^{-3/2} M_g^{-1/2} \quad (3)$$

by optically thick synchrotron radiation as it spiralled in from r to $r/2$ (where $\dot{m}_{-4} = 10^4 \dot{m}$ and $\alpha_{-2} = 10^2 \alpha$). The actual cooling rate is considerably higher for

$$r < r_{\text{comp}} \equiv \left(\frac{m_p}{m_e} \right)^{4/5} (\dot{m} \alpha^{-1})^{2/5} = 65 \dot{m}_{-4}^{2/5} \alpha_{-2}^{-2/5} \quad (4)$$

since then the dominant cooling is due to Compton scattering of the synchrotron photons (Fig. 1). Thus if the electrons and ions are kept in energy equipartition, a thick torus will cool and deflate to a thin disk for $r \leq 50 r_g$.

The torus will be 'collisionless' inside some large radius (typically $\sim 2,000 r_g$, but defined more precisely below), in the sense that through two-body collisions alone protons at temperature T_p give less than half their energy to electrons (with $T_e < T_p$) on the inflow time scale, if

$$\dot{m} < \left(\frac{m_p}{m_e} \right) (\ln \Lambda)^{-1} \alpha^2 \left(\frac{m_p}{m_e} \frac{r_g}{r} \frac{T_e}{T_p} \right)^{3/2} \approx 50 \alpha^2 \left(\left(\frac{2,000 r_g}{r} \right) \left(\frac{T_e}{T_p} \right) \right)^{3/2} k T_e < \frac{2}{3} m_e c^2 \quad (5a)$$

$$\dot{m} < \left(\frac{m_p}{m_e} \right) \left(\frac{\gamma_{\text{re}}}{\ln \Lambda} \right) \alpha^2 = 50 \gamma_{\text{re}} \alpha^2 \quad k T_e > \frac{2}{3} m_e c^2 \quad (5b)$$

In these conditions, collective plasma phenomena will determine the particle distribution functions, the radiative efficiency, and thus the pressure and shape of the whole torus/disk. In most of what follows we assume that collective processes are unable to transfer more than half the proton's kinetic energy to the electrons [which would in turn quickly radiate it: compare with equation (3)]. Given the complexity of the plasma physics and the uncertainties in accretion theory, it may be a long time before this assumption can be checked. If the collective transfer can take place, the structure is similar to that already familiar from results at high accretion rates. If it cannot, one consequence, as already realized in the context of accretion disks around stellar-mass black holes^{35,36}, is a 'two-temperature' flow.

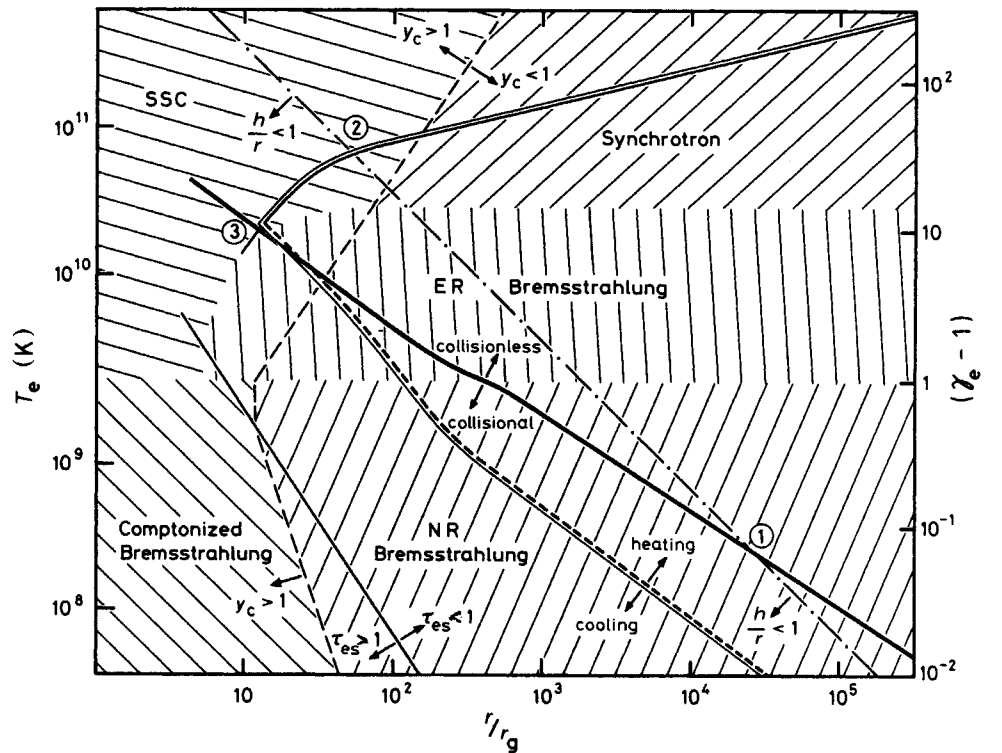
Two-temperature tori

Consider an accretion flow where equation (5) is satisfied for $r < r_D$, and assume that the protons and electrons are then

Fig. 1 The dominant local cooling mechanisms for a representative ($\alpha = 10^{-2}$, $\beta = 1$) low-accretion rate ($\dot{m} = \dot{M}/\dot{M}_{\text{Edd}} = 10^{-4}$) flow about a central mass of $10^8 M_\odot$, as a function of distance (measured in units of $r_g = GM/c^2$) from the central mass and electron temperature (which also parametrizes disk thickness; a scale of electron kinetic energy in units of $m_e c^2$ is at right). We have assumed electron and ion temperature to be everywhere equal, although Coulomb collisions are insufficient to maintain equipartition for disk states above the heavy solid line. In the C-shaped region to the left of the double lines the cooling rate exceeds the rate of viscous heating and the disk will collapse. To the right of the double lines the rate of heating exceeds that of cooling and the disk will thicken (states above the $h/r = 1$ line are unphysical in the absence of external pressure forces). SSC denotes synchrotron self-Compton cooling ('the Compton catastrophe'), τ_{es} is the optical depth to Thomson scattering

$$\gamma_c = \begin{cases} \frac{3}{2}(\gamma_e - 1)\tau_{\text{es}}(\tau_{\text{es}} + 1) & \text{for } \gamma_e - 1 \ll 1 \\ \frac{1}{2}\gamma_e^2 \tau_{\text{es}}(\tau_{\text{es}} + 1) & \text{for } \gamma_e - 1 > 1 \end{cases}$$

is the Compton y -parameter. Not included are: induced Compton cooling, non-local Compton cooling/heating and losses to winds. The disk is everywhere supported by gas pressure. Radiation pressure is only important in the inner parts of cold, thin disks for $\dot{m} > 10^{-3}$. If, for $r \leq 10^7 r_g = 100 M_g$ pc, the disk is heated to a temperature above that at which bremsstrahlung cooling becomes inefficient (double dashed line) it will thicken to a torus, $h \sim r$. For $r < r_D = 2 \times 10^4 r_g$ (point 1), Coulomb collisions cannot maintain equipartition. If plasma processes take over this job, the disk will begin to cool by SSC at $r \approx 50 r_g$ (point 2) and will gradually thin to $(h/r) \sim \frac{1}{2}$ at $r = 10 r_g$, at which point (3) bremsstrahlung will rapidly cool it to a new equilibrium at $T \sim 10^4$ K and $(h/r) \sim 10^{-4}$ where it radiates as a black body (optically thick bremsstrahlung and line cooling just balance viscous heating). If, as assumed here, equipartition cannot be maintained, the electrons continue to cool (Fig. 3) inside $r = r_D$ (point 1) but the ions remain at the virial temperature. The torus will thus remain thick, with the structure shown in Fig. 2.



indeed thermally decoupled from each other. Equation (2) shows that, for $\dot{m}_{-4} \approx 1$, the thick ($h \approx r$) torus is a self-consistent solution even for $r \gg r_R = 2,000 r_g$ (see Fig. 1). At $r_D \gg r_R$, electrons and ions will decouple; but as neither can cool, they may remain at the same temperature. For $r < r_R$, the electrons become relativistic. If the heating of both species is adiabatic, then $T_p \propto n^{2/3}$ while $T_e \propto n^{1/3}$, so the ions are preferentially heated. Shock heating may likewise favour the ions, as isotropization of the bulk flow velocities will give to each species a thermal energy proportional to its mass. Fermi acceleration processes will amplify any temperature difference once it exists. Thus the electrons are granted little of the ions' energy, despite their ability to radiate it [equation (3)]. The ions remain at the virial temperature, and the two-temperature torus remains thick even inside $r \approx 50 r_g$ [equations (3) and (4)].

If the torus surrounds a spinning black hole, the Lense-Thirring effect can enforce axisymmetry with respect to the hole's spin axis³⁷ out to a radius $\sim \alpha^{-2/3} (J/J_{\max})^{2/3} r_g$, where J is the hole's angular momentum and $J_{\max} = M r_g c$ is the maximum value allowed by the Kerr metric. We would expect the holes to have been spun up to $J \approx J_{\max}$ during their previous quasar phase^{38,39}. There may be misalignment outside this radius if the gas is not supplied in the equatorial plane of the hole. The isotherms within the torus can then be calculated in the same way as for radiation-supported tori^{18,19,22}. If the pressure support is due primarily to non-relativistic ions, then the effective specific heat ratio will be about 5/3. If α is independent of radius, then $n \propto r^{-3/2}$, and most of the mass of the torus resides at large radii. If both these conditions are fulfilled, then the ions are roughly isentropic.

The contours of Fig. 2 have been computed by this procedure. The results are subject to the same uncertainties as those for radiation-supported tori: problems with self-consistency in viscosity (R.D.B. and M. Jaroszyński, in preparation), and the absence of a formalism for properly matching one of these inner solutions onto a flow field specified at large radii. Magnetic fields will modify the simple structure shown. However, essential features, such as the characteristic vortex funnel, which becomes parabolic in the limit of a flow with constant specific angular momentum l and is less collimated than parabolic for a flow with l increasing outwards, are unlikely to be changed. (Flows for which l decreases outwards are argued to be unstable; note, however, that l must be constant along the surface if there is no poloidal circulation and we insist that the surfaces have zero energy-at-infinity.) In this 'zone of non-stationarity', matter cannot have enough angular momentum to achieve near-equilibrium between centrifugal force, pressure and gravity. A fluid flow in the funnel must either fall straight into the hole, or have enough enthalpy to leave the region rapidly. Hydromagnetic effects⁴⁰ or flares (compare with previous remarks about buoyant fields) on the funnel walls will inject into it particles of positive energy-at-infinity. The liberated power will escape along the two funnels, leading to a pair of inertially collimated jets carrying some fraction of $\dot{M} c^2 = 10^{42} \dot{m}_{-4} M_8 \text{ erg s}^{-1}$. This may be enough to explain the low-luminosity double-jet radio sources.

Electromagnetic extraction of the hole's rotational energy

If the central black hole is near-maximally rotating ($J \approx J_{\max}$), a much greater supply of free energy is available: the rotational energy of the hole (up to 29% of $\dot{M} c^2$ for $J = J_{\max}$). If an ordered electromagnetic field can be supported around the hole, then it can put particles (which may in fact be inside the horizon if the field threads the latter) systematically on negative energy orbits. The negative energy swallowed by the hole appears as positive energy in the external fields and thus drives away from the hole a Poynting flux of luminosity⁴¹⁻⁴³

$$L_{\text{EM}} \leq B_{\text{pH}}^2 r_g^2 (J/J_{\max})^2 c \quad (6)$$

where B_{pH} is the poloidal field in the flux tube that intersects the

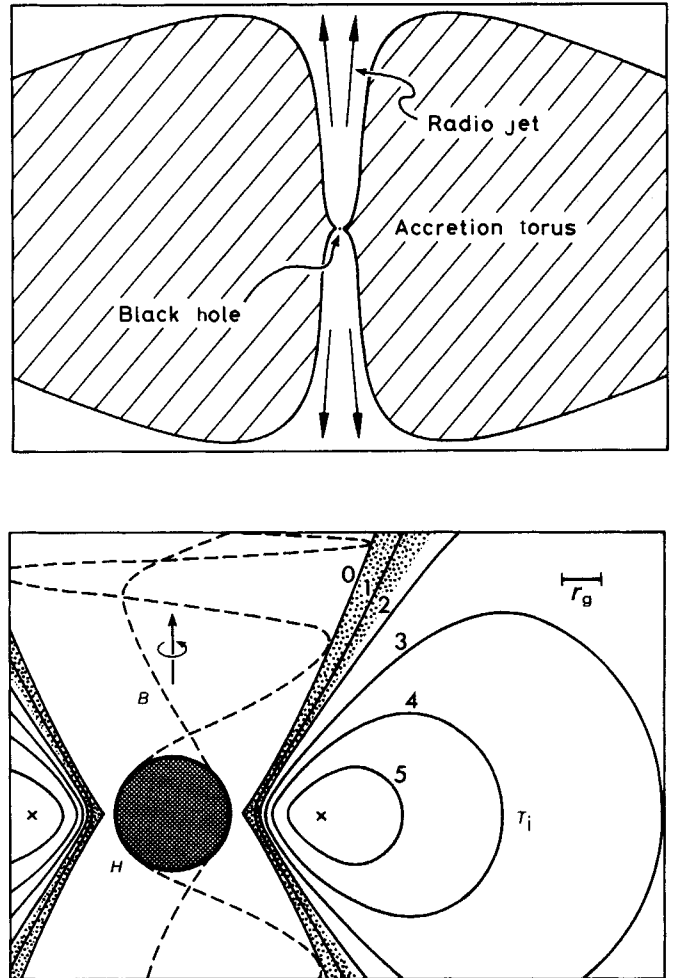


Fig. 2 An accretion torus supported by ion pressure surrounds a spinning black hole. The torus defines a pair of funnels capable of collimating a relativistic hydromagnetic flow which ultimately forms the observed radio jets. The enlargement shows the inner regions of the torus and displays contours of ion temperature in units of 10^{11} K (solid lines). These have been calculated assuming that the ions are an isentropic gas with constant specific angular momentum and that the hole (horizon H) has $J = 0.9 J_{\max}$ (see ref. 19). The outer isotherm is also the surface of zero binding energy. The dashed lines represent the magnetic field lines in the configuration discussed in the text. A poloidal current flows in the funnel. Large toroidal currents and the return poloidal currents flow in a thin surface layer (stippled).

hole's horizon. If $J = J_{\max}$, then this power $\leq 10^{44} M_8 \alpha^{-1} \dot{m}_{-4} \beta_{\text{pH}}^{-1} \text{ erg s}^{-1}$ (where $\beta_{\text{pH}} = (\text{maximum gas pressure in torus}) / (B_{\text{pH}}^2 / 8\pi)$) can exceed by $\approx \alpha^{-1}$ the rate of liberation of binding energy by the accreting matter. The torus will be able to confine and initially collimate the extracted power, as the associated radiation (or ram) pressure will be less than the gas pressure in the torus if

$$\dot{M} c^2 / 4\alpha > L_{\text{EM}} \leq \dot{M} c^2 / (20\alpha\beta_{\text{pH}}) \cdot (J/J_{\max})^2 \quad (7)$$

Both these inequalities are nearly always satisfied since $\beta_{\text{pH}} \gg 1$.

To extract power predominantly from the hole, we must have $\beta_{\text{pH}} \approx 1$ and $\alpha \ll 1$. Is this consistent with a physically attainable disk structure? One might naively suppose that the poloidal field could be generated by currents distributed through the torus, and thus be an extension of an ordered field structure within it. But if this were the case, the internal fields would transport energy and angular momentum outwards so efficiently that $\alpha \approx 1$ and L_{EM} would be $\leq \dot{M} c^2$. Furthermore, in the presence of a magnetosphere any current loop in keplerian rotation around a central mass loses energy from its outer light cylinder at a rate ($\sim$ vacuum magnetic dipole power) $\approx (J_{\max}/J)^2$ times the rate at

which it could extract power were the central mass a black hole of angular momentum J . If the current loop runs in an accretion disk, the missing power must be provided by the liberation of binding energy, which again implies $L_{EM} \approx \dot{M}c^2$ (compare with the thin-disk solution of Blandford and Znajek⁴² in which $L_{EM} \approx 0.3 (J/J_{\max})^2 L_{\text{disk}}$).

To understand how a torus could avoid all this energy transport and loss, it is helpful to consider the time evolution of a perfectly conducting torus with a very weak ($\beta_p \gg 1$) magnetic field piercing it at large radii. (It may also have internally closed field loops of arbitrary strength.) This weak field is convected in with the flow, and by applying Faraday's law to a sufficiently large loop encircling the rotation axis, we find that the magnetic flux through the funnel increases as we carry field in, even though the field lines are causally disconnected from the matter in which they were frozen after the latter has crossed the horizon⁴³. If the torus behaves like a perfect conductor, it will exclude this increasing external flux by developing surface currents. The field and surface currents will increase (perhaps shutting off the accretion altogether) until either some massive plasma-field interchange instability leads to an explosive energy release or, more likely, dissipation in the surface layer destroys piercing field as fast as it is convected in. The structure of such a torus is rather different from that of the field-free torus: the $\mathbf{J}_T \times \mathbf{B}_p$ force across the surface layer (funnel boundary) must be balanced by a gradient of gas pressure, so that at its base $P_{\text{gas}} \approx B_{\text{pH}}^2/8\pi$. If we want $\beta_{\text{pH}} \approx 1$, most of the drop of gas pressure must occur in the surface layer, the inner parts of the torus being at roughly constant pressure.

We can imagine the following self-consistent solution for a torus with a jet (Fig. 2): toroidal currents in the thin surface layer of the vortex funnel produce a poloidal field which is torqued by the hole, driving an outgoing Poynting flux. This will soon become plasma-loaded; toroidal fields will begin to dominate, causing self-collimation^{34,44,45}. The return poloidal current must also flow in the thin surface layer, pulling material from it into the funnel, but leaving the bulk of the torus unaffected. The only forces from the jet transmitted to the body of the torus are those due to the poloidal field (transmitted by the gas pressure in the subcutis), whose pressure drops as (area of funnel)⁻² and which thus has less effect on the structure of the torus the further out we go. Simple solutions ignoring the jet can therefore correctly model the torus, while the jet may collimate itself on the paraboloidal field lines thus defined without exerting much modifying force.

The above discussion shows how a torus supported by ion pressure, with a low accretion rate $\dot{M}$, can catalyse the extraction of a hole's spin energy at a rate $\approx \dot{M}c^2$. Moreover, we show below that the luminosity from the torus itself may be $\ll \dot{M}c^2$. Thus if the meagre losses from the torus can be replaced by energy and angular momentum extracted from the hole, there need be no accretion: only the inertia of the torus, and the currents in it are needed to confine the magnetic fields that tap the hole's energy.

Losses and radiation from the torus

The total power dissipated in the surface layer of thickness s is $\sim (s/r_g)L_{EM} \ll L_{EM}$. Moreover, as only a very weak field need penetrate the torus, transport and losses from the outer light cylinder may also be negligible. The radiation emitted by the body of an ion-supported torus will be primarily non-thermal; its magnitude depends sensitively on the efficiency of electron acceleration. The torus will generally be optically thin to Thomson scattering, with a total optical depth along a line of sight in the equatorial plane, from the hole to infinity, of $\sim \dot{m}\alpha^{-1}$. Incoherent cyclotron and synchrotron radiation will be self-absorbed. (The conditions envisaged resemble those of a so-called 'plasma turbulent reactor'⁴⁶, so that a power law spectrum may be set up). The torus will, however, be optically thin to γ radiation produced by relativistic electron bremsstrahlung, and to inverse Compton radiation at frequencies above 10^{13} Hz. The

relevant relaxation times that go into fixing the electron temperature are plotted in Fig. 3, which shows that the electrons will be able to radiate efficiently from the inner part of the torus only if their temperatures exceed 10^{10} K. At higher temperatures the power radiated increases very rapidly with temperature: the optically thick synchrotron losses vary as T_e^7 , and the dependence on T_e is steepened further because repeated Compton scattering of synchrotron photons up to the γ -ray band dominates when $\gamma_e^2\tau_{es} > 1$. However, as long as equation (5) is satisfied, the electrons can radiate only the small fraction of the total energy which they can drain from the ions (or derive from adiabatic heating) and the ion-supported torus will not collapse. For the parameter range of interest, the radiation will be in the far IR or the γ -ray region of the spectrum, according as $\dot{m} \lesseqgtr 10^{-5}\alpha^{-3/5}\beta^{-3/5}M_8^{-1/5}$. We cannot exclude the possibility that coherent radiation from fields ~ 100 G may be important. But, although this might seem relevant to the phenomenon of low-frequency radio variability⁴⁷ it would be subject to electron synchrotron absorption and induced Compton scattering^{48,49}.

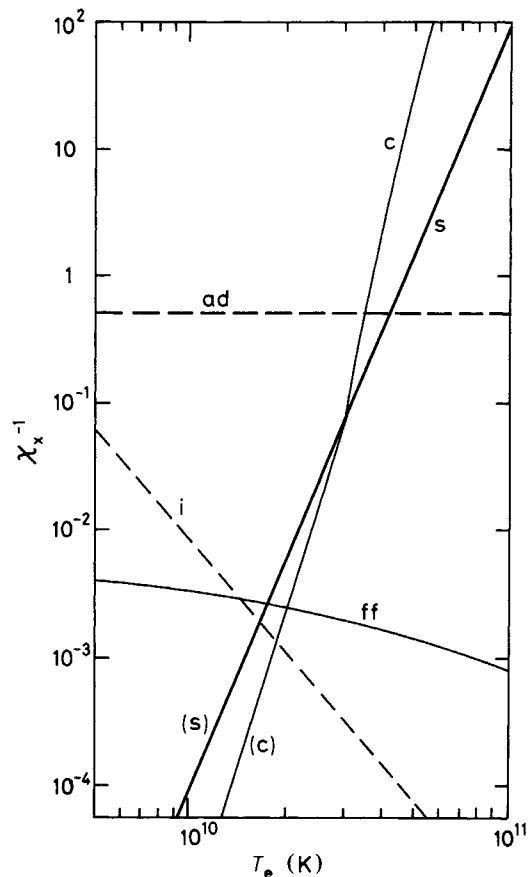


Fig. 3 Dimensionless electron relaxation rates χ_x^{-1} as a function of the electron temperature T_e . χ_x^{-1} is related to the heating or cooling rate per unit volume ϵ_x by

$$\chi_x^{-1} = \frac{\epsilon_x}{3n^2kT_e\sigma_c}$$

The lines are drawn for $\alpha = 10^{-2}$, $\beta = 8\pi nkT_i/B^2 = 1$, $\dot{m} \equiv \dot{M}/\dot{M}_{\text{Edd}} = 10^{-4}$, $r_*/r_g = 5$, and $M = 10^8 M_\odot$. The relaxation rates with appropriate scalings are: compressional heating, $\chi_x^{-1} \propto \alpha^2 \dot{m}^{-1}$; collisional heating by ions, $\chi_x^{-1} \propto r_*^{-1}$; optically thick synchrotron cooling, $\chi_x^{-1} \propto \alpha^{1/2} \beta^{-5/2} r_*^{-7/4} \dot{m}^{-1/2} M^{-1/2}$; optically thin Compton cooling, $\chi_x^{-1} \propto \alpha^{-2} \beta^{-1} r_*^{-3.25} \dot{m}^2$, and relativistic electron bremsstrahlung χ_x^{-1} (no scaling). Self-absorption of the Compton radiation, double-Compton and pion cooling and pair production are unimportant for the range of densities and temperatures under consideration. The virial temperature of the ions is 7×10^{11} K ($r_*/5$)⁻¹ (see also Fig. 2), while in the absence of losses, compressional heating of electrons transrelativistic at $r_* \approx 1,000$ would give $T_e \approx 3 \times 10^{10} (r_*/5)^{-1/2}$ K.

The γ -ray flux from the inner regions of the torus (which may include a redshifted positron annihilation line) would have only a small probability of being reabsorbed near the hole. Nevertheless, the small number of interactions that will occur can produce sufficient electron-positron pairs within the funnel to complete the circuit and allow electric current to flow through the hole. For example, to support an electromagnetic power of $10^{44} \text{ erg s}^{-1}$ in the funnel, a current of $3 \times 10^{17} \text{ A}$ must flow, which requires a minimum pair creation rate of $\sim 10^{36} \text{ s}^{-1}$. This can be maintained by γ rays from the torus.

If, as assumed above, the electrons and ions are indeed thermally decoupled when equation (5) is satisfied, then the inner part of the torus will radiate much less than $0.1 \dot{M} c^2$, and this power is in the far IR or the γ -ray band. But if, on the other hand, collective plasma effects couple the ions and electrons at all radii, then (see Fig. 1) there will be a thin disk for $r \leq 50 r_g$, radiating $\sim 0.1 \dot{M} c^2$ with a spectrum peaking in the optical. For $\dot{m} > 50 \alpha^{-2}$, even Coulomb coupling of electrons and ions is efficient enough to cool the ions, and one would expect a thin disk radiating in the near UV. When $\dot{m}$ is low, the predicted emission from galactic nuclei depends so drastically on whether or not the electrons and ions are coupled by plasma effects, that this issue can perhaps be settled observationally more readily than by the efforts of plasma theorists.

The maximum power that can be self-consistently extracted from a black hole by a thick ion-supported torus, $\dot{M}$ being given, is achieved when α has the lowest value consistent with the electron-ion decoupling condition $\dot{m} < 50 \alpha^2$. Even if $\dot{M}$ is a free parameter, the maximum power that can be self-consistently extracted from a black hole by a thick torus supported by ion pressure is $\sim 10^{45} \alpha^{-2} \dot{M}_8 \text{ erg s}^{-1}$. However, if strong magnetic fields are anchored in a dense radiation-pressure-supported torus¹⁸⁻²¹, whose thermal luminosity is $\geq L_{\text{Edd}}$, there is no clearcut reason why the power extracted electromagnetically from the hole should not be orders of magnitude larger. A model along these lines may be relevant to objects such as 3C273 where (in contrast to the radio galaxies) there is a nuclear thermal luminosity $\sim L_{\text{Edd}}$, but where there is evidence also for a powerful relativistic jet.

Astrophysical context

We have argued that, provided the viscosity parameter α is not too low and the electron-ion coupling is sufficiently weak, a two-temperature, ion-supported torus can be maintained around a massive black hole accreting gas at a very subcritical rate in a galactic nucleus. The inner region of such a torus can collimate a pair of relativistic jets comprising electrons, posi-

trons and electromagnetic fields. If radio jets originate in the funnels of accretion tori, the latter are likely to be supported by ion pressure rather than radiation pressure in most observed cases. The luminosity of the torus would then be far less than the kinetic energy flux in the jets. (The scale of the observed jets is, of course, many orders of magnitude larger than the torus itself, so some further collimation process may also be operative).

If most of the power is extracted from a spinning black hole, a fuelling rate of $\sim 10^{-3} \dot{M}_8 \text{ yr}^{-1}$ suffices for most radio sources. A 'baseline' value for $\dot{M}$ in galactic nuclei would be supplied by debris from stars which pass so close to the hole that they are tidally disrupted^{4,32,33}. This debris would have an angular momentum appropriate to an orbit at a few times r_g (for $M_8 \approx 1$) but a specific binding energy of only $\sim 10^{-5} c^2$ (ref. 33). This debris would then form a torus extending out to $\sim 10^5 r_g$; a single disrupted star could then maintain $\dot{m} \approx 10^{-4}$ for $\sim 10^4 \text{ yr}$.

If the energy in radio source jets derives mainly from the spin of the central black hole, then the hole probably underwent a phase of rapid mass accretion at an earlier epoch. One can speculate on a possible evolutionary sequence governed by accretion rate. If $\dot{m} > 10$, a radiation-supported torus can be formed. For $50 \alpha^2 \leq \dot{m} \leq 10$ the situation is less clear, but an ion-supported torus could possibly result, with most of the mass being contained within cool clouds or a thin disk.

Disks with $\dot{m} > 50 \alpha^2$ will be strong sources of ionizing radiation, and should be associated with strong line emission from the nucleus. Tori with $\dot{m} < 50 \alpha^2$ will be essentially invisible except as γ -ray sources. Both may drive the strongest of extended radio sources. Recent quasar observations have shown that the majority are radio-quiet, have small percentage polarization, and relatively low X-ray luminosity⁵⁰⁻⁵². Presumably these do not create large fluxes of relativistic particles, and are perhaps $\sim 10^8 M_\odot$ black holes surrounded by radiation-supported tori without dominant magnetic fields. Seyfert galaxies may be similar, but with $\sim 10^6 M_\odot$ black holes. The radio loud and (usually) X-ray luminous objects may then be associated with electromagnetically powered tori, the BL Lac and superluminal radio sources perhaps being viewed at small angles to the jet axis.

Finally, the physics we have been discussing, with the exception of the detailed radiative properties of the torus, involves no intrinsic mass scale. Indeed, there may be an ion-supported torus around a $\leq 10^6 M_\odot$ hole in our galactic centre, and relativistic jets collimated by tori around stellar-mass black holes may exist within our Galaxy.

M.C.B. acknowledges partial support under NSF grant AST 79-23243. R.D.B. acknowledges hospitality at the Institute of Astronomy and support under NSF grant AST 80-11752 and from the Alfred P. Sloan foundation. E.S.P. thanks the Marshall Commission for financial support.

Received 28 July; accepted 9 November 1981.

- Wyckoff, S., Wehinger, P. & Gehren, T. *Astrophys. J.* **247**, 750 (1981).
- Schmidt, M. *Phys. Scr.* **17**, 135 (1978).
- Soltan, A. *Mon. Not. R. astr. Soc.* (in the press).
- Frank, J. *Mon. Not. R. astr. Soc.* **184**, 87 (1978).
- Gunn, J. E. *Active Galactic Nuclei* (eds Hazard, C. & Mitton, S.) (Cambridge University Press, 1978).
- Sparke, L. S. & Shu, F. H. *Astrophys. J. Lett.* **241**, L65 (1980).
- Barr, P. et al. *Mon. Not. R. astr. Soc.*, **193**, 549 (1980).
- Osterbrock, D. E. & Miller, J. S. *Astrophys. J.* **197**, 535 (1975).
- Koski, A. T. *Astrophys. J.* **223**, 56 (1978).
- Costero, R. & Osterbrock, D. E. *Astrophys. J.* **211**, 675 (1977).
- Waggett, P. C., Warner, P. J. & Baldwin, J. E. *Mon. Not. R. astr. Soc.* **181**, 465 (1977).
- Cohen, M. H. & Readhead, A. C. S. *Astrophys. J. Lett.* **233**, L101 (1979).
- Readhead, A. C. S. et al. *Nature* **276**, 768 (1978).
- Linfield, R. P. *Astrophys. J.* **244**, 436 (1981).
- Pearson, T. J. et al. *Nature* **290**, 365 (1981).
- Cotton, W. D., Shapiro, I. I. & Wittels, J. J. *Astrophys. J. Lett.* **244**, L57 (1981).
- Jones, D. L., Sramek, R. A. & Terzian, V. *Astrophys. J. Lett.* **247**, L57 (1981).
- Abramowicz, M. A., Jaroszynski, M. & Sikora, M. *Astr. Astrophys.* **63**, 221 (1978).
- Kozłowski, M., Jaroszynski, M. & Abramowicz, M. A. *Astr. Astrophys.* **63**, 209 (1978).
- Abramowicz, M. A., Calvani, M. & Nobili, L. *Astrophys. J.* **242**, 772 (1980).
- Paczynski, B. & Wiita, P. *Astr. Astrophys.* **88**, 23 (1980).
- Jaroszynski, M., Abramowicz, M. A. & Paczynski, B. *Acta Astr.* **30**, 1 (1980).
- Abramowicz, M. & Piran, T. *Astrophys. J. Lett.* **241**, L7 (1980).
- Sikora, M. *Mon. Not. R. astr. Soc.* **196**, 257 (1981).
- Sikora, M. & Wilson, D. B. *Mon. Not. R. astr. Soc.* (in the press).

- Eardley, D. M. & Lightman, A. P. *Astrophys. J.* **200**, 187 (1975).
- Shakura, N. I. & Sunyaev, R. A. *Astr. Astrophys.* **24**, 337 (1973).
- Novikov, I. D. & Thorne, K. S. in *Black Holes* (eds DeWitt, C. & B.S.) (Gordon & Breach, New York, 1973).
- Pringle, J. E. *Ann. Rev. Astr. Astrophys.* **19**, 137 (1981).
- Pringle, J. E. *Mon. Not. R. astr. Soc.* **177**, 65 (1976).
- Shakura, N. I. & Sunyaev, R. A. *Mon. Not. R. astr. Soc.* **175**, 613 (1976).
- Hills, J. G. *Mon. Not. R. astr. Soc.* **182**, 517 (1978).
- Frank, J. *Mon. Not. R. astr. Soc.* **187**, 883 (1979).
- Blandford, R. D. & Payne, D. *Mon. Not. R. astr. Soc.* (in the press).
- Shapiro, S. L., Lightman, A. P. & Eardley, D. M. *Astrophys. J.* **204**, 187 (1976).
- Eardley, D. M., Lightman, A. P., Payne, D. G. & Shapiro, S. L. *Astrophys. J.* **224**, 53 (1978).
- Bardeen, J. M. & Petterson, J. A. *Astrophys. J. Lett.* **195**, L65 (1975).
- Thorne, K. S. *Astrophys. J.* **191**, 507 (1974).
- Abramowicz, M. A. & Lasota, J. P. *Acta Astr.* **30**, 35 (1980).
- Galeev, A. A., Rosner, R. & Vaiana, G. S. *Astrophys. J.* **229**, 318 (1979).
- Znajek, R. L. thesis. Cambridge Univ. (1976).
- Blandford, R. D. & Znajek, R. L. *Mon. Not. R. astr. Soc.* **179**, 433 (1977).
- Macdonald, D. & Thorne, K. S. *Mon. Not. R. astr. Soc.* (in the press).
- Blandford, R. D. *Mon. Not. R. astr. Soc.* **176**, 465 (1976).
- Chan, K. L. & Henriksen, R. N. *Astrophys. J.* **241**, 534 (1980).
- Norman, C. A. & ter Haar, D. *Phys. Rep.* **17**, 307-358 (1975).
- Condon, J. J. et al. *Astr. J.* **84**, 1 (1979).
- Levich, E. V. & Sunyaev, R. A. *Astrophys. Lett.* **7**, 69 (1970).
- Wilson, D. B. & Rees, M. J. *Mon. Not. R. astr. Soc.* **185**, 297 (1978).
- Condon, J. J. et al. *Astrophys. J.* **244**, 5 (1981).
- Ku, W. H. M., Helfand, D. J. & Lucy, L. B. *Nature* **288**, 323 (1980).
- Angel, J. R. P. & Stockman, H. S. A. *Rev. Astr. Astrophys.* **18**, 321 (1980).

Structural evolution of a crustal section in the western Himalaya

M. P. Coward*, M. Q. Jan†, D. Rex*, J. Tarney‡,
M. Thirlwall* & B. F. Windley‡

* Department of Earth Sciences, The University, Leeds LS2 9JT, UK

† Department of Geology, University of Peshawar, Peshawar, Pakistan

‡ Department of Geology, The University, Leicester LE1 7RH, UK

Recent field work on the Kohistan complex in the Karakorum Range of North Pakistan has shown its stratigraphy to be repeated by several phases of tight to isoclinal folding. This reveals a close association between igneous activity and a late deformation and complicates the previously proposed island arc model.

AN exciting crustal section has recently been discovered in the Karakorum Range of North Pakistan^{1,2}. This was considered to be a complete section of an island arc (the Kohistan complex or sequence) thrust over rocks of the Indian plate and turned on end during Himalayan orogeny (for details of the regional setting, see ref. 3). The arc was initially interpreted as a stratigraphic sequence from upper crustal sediments and volcanics in the north through tonalites and diorites to lower crustal granulites in the south. Our recent field work has shown a more complex structure, stratigraphy being repeated by several phases of tight to isoclinal folding. The Kohistan sequence forms a major tight syncline whose plunge varies over 90° and this folds a complete section through the crust, the lower crustal rocks having previously been isoclinally folded. The Kohistan sequence was thrust southwards over the gneisses of the Indian plate and the earlier arc structures were steepened to a near-vertical attitude. Most of the tonalite–diorite suite was intruded during this deformation sequence, but some outlasted the deformation. This close association between igneous activity and late deformation complicates a simple arc model.

North–south traverse

Figures 1 and 2 show a simplified map and section of the Kohistan area adapted from refs 3, 4 and a new unpublished map by R. A. K. Tahirkheli. We consider here a north to south section from the Karakorum Ranges to the molasse deposits of the Indian plate. In the Hunza valley in the north there is a chaotic arrangement of lithologies including limestones, sandstones, conglomerates, basic volcanics and ultrabasic rocks such as serpentinites, meta-harzburgites, talc-carbonate and talc-chlorite schists, all occurring as lenses several kilometres long separated by chloritoid slates. These slates increase in grade rapidly northwards to garnet–staurolite mica schists intruded by the granodiorites of the Karakorum batholith. The slates and schists contain a well developed schistosity folded around a large-scale, gently dipping synform (see Fig. 2). This chaotic pile of lithologies resembles a large-scale tectonic mélange which is situated along the northern suture.

To the south there are gabbros, basic volcanics, including spectacular sections of deformed pillow lavas, volcanoclastic sediments and greywackes folded into tight large-scale anticlines and synclines. Further south these sediments and volcanics are invaded by tonalites, diorites and large pegmatite swarms, locally constituting 30% of the rock volume. Between these intrusions there are screens of graded greywackes and pillow lavas from which younging directions can be obtained (see Fig. 1). This sequence extends for ~70 km from the northern suture. The dominant structures are large-scale, upright antiforms and synforms, whose hinges plunge steeply NNE parallel to mineral lineations down the dip direction of the cleavage. The major structure shown on the map is a westward-

facing, steeply plunging anticline at Gilgit and a complementary major syncline with a half wavelength of >50 km together with common smaller-scale parasitic folds. These structures bring the major stratigraphy from north of Gilgit southwards to Jaglot, particularly repeating the spectacular pillow lavas seen in the north. This same Jaglot syncline probably continues westwards to the central part of Kohistan where it contains the Eocene Kalam volcanics⁵ and underlying sediments; in this region the syncline has a more horizontal plunge.

On the south limb of the syncline one passes down into a vast stratiform cumulate complex containing chromite-layered dunites, troctolites, gabbros and predominant norites with spectacular phase graded units. There are very spectacular and abundant graded units which have allowed numerous younging directions to be obtained. In the Chilas area these layers are deformed into large-scale steeply plunging folds, parasitic on the Jaglot syncline. However, there are changes in facing direction as shown by the cumulate grading indicating an earlier major isoclinal fold phase. The central part of the Chilas complex is composed of a zone of ultramafic rock and this youngs to metanorite at the northern and southern margins. The whole Chilas complex is interpreted as one isoclinal anticline refolded by Jaglot syncline age structures. The complex has suffered metamorphism of two pyroxene granulite facies, the minerals of which are aligned in the fabric of Jaglot syncline age.

There are two important later stages of deformation which affect both the Chilas complex and the metasediments and igneous rocks to the north. A phase of southward dipping, small-scale folds produces crenulation cleavage in the more schistose rocks, possibly related to the nearly recumbent fold in the Hunza valley. Later deformation produced a major conjugate set of NW–SE dextral and NE–SW sinistral trending shear zones with dominantly strike-slip sense of displacement: these vary from ductile to brittle, and from phyllonites of amphibolite–greenschist facies to pseudotachylites.

To the south of the Chilas complex there is a complex sequence of metasediments, amphibolites, gabbros, deformed granites and pegmatites and post-tectonic granites. Most rocks show two phases of deformation: the dominant second phase deforms early fabrics and folds, and produces steeply-plunging folds parasitic on the Jaglot syncline as shown in Fig. 2.

At Patan there is the Jijal complex⁶ of chromite-layered dunites and granulites, which we interpret as part of the Chilas complex metamorphosed to a high-pressure (garnet–clinopyroxene) granulite grade. This interpretation requires a first-phase tight syncline between the Chilas and Jijal complexes. The Jijal complex shows two phases of deformation similar to those described above, the first fabric of which is difficult to detect, being largely overgrown by granulite facies assemblages. To the south the rocks become intensely sheared and metamorphosed to the amphibolite facies. All these structures become more

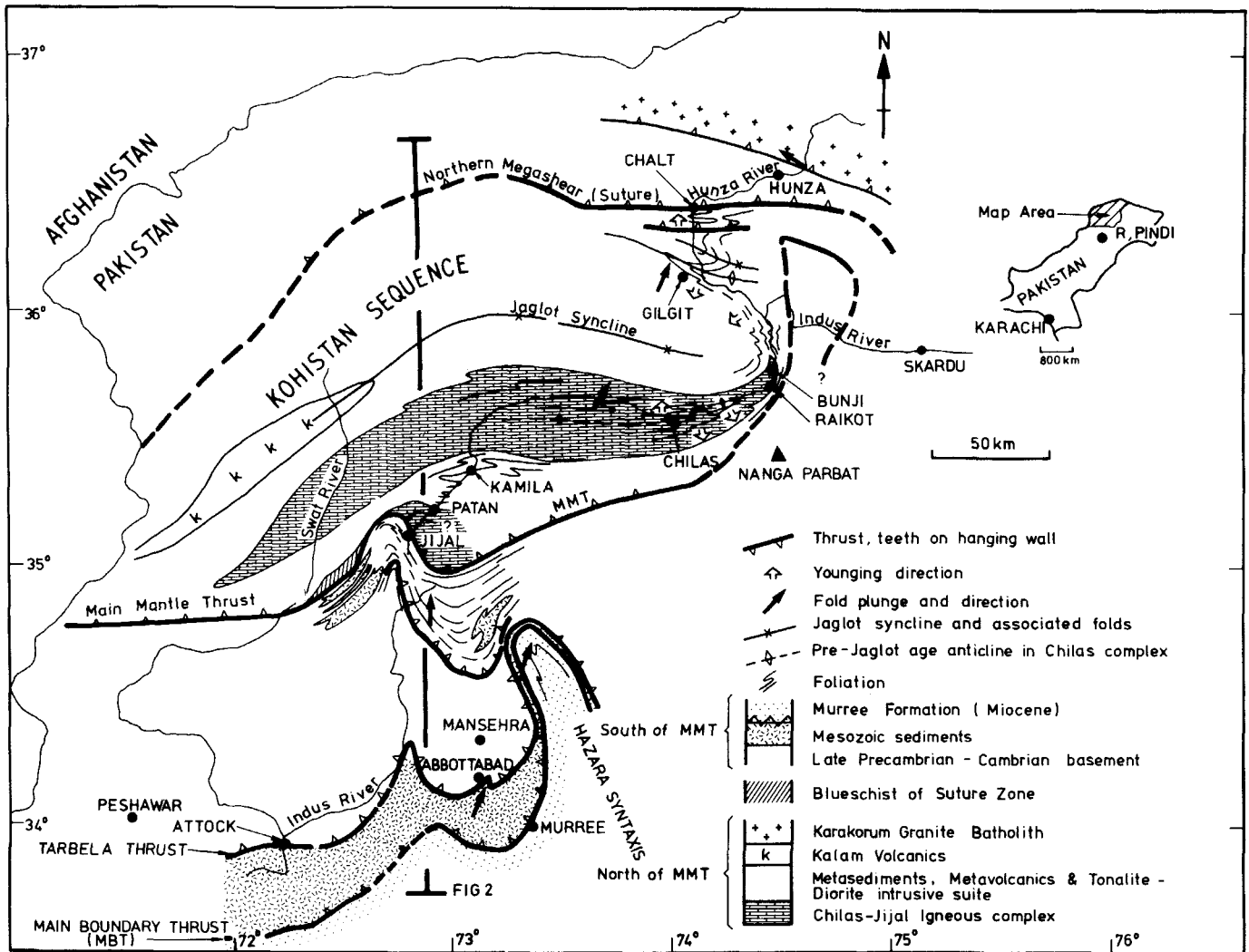


Fig. 1 Map of the Kohistan region showing the main structures (modified from refs 2 and 4).

gently dipping to the south, so that near Jijal the fabrics are flat-lying or dip only gently northwards. Along strike to the south-west there is a 10–15 km wide thrust wedge of blueschists⁷ along the Main Mantle Thrust (MMT) (Fig. 1), though to the east, around the Nanga Parbat bulge (Fig. 1), no blueschists occur, the MMT is difficult to detect and the obvious faults and shears are late structures which post-date the folding.

At Jijal a major shear zone, the MMT^{1,2}, carries the Jijal complex and the Kohistan sequence southwards over quartzofeldspathic gneisses of the Indian plate containing an intense

fabric folded by recumbent, nearly isoclinal folds with hinges plunging to the NNE. Both sets of structures are folded around large-scale upright F_3 folds. The gneisses with intense deformation and isoclinal folds can be followed for 10–20 km south of the MMT and contain infolded synclines of Mesozoic cover rocks (marbles, graphite schists and conglomerate). Further south where deformation decreases, the unconformity can be seen between these sediments and the underlying basement gneiss. Where least deformed by Himalayan shearing, as at Mansehra, this basement appears as a homogeneous porphyritic

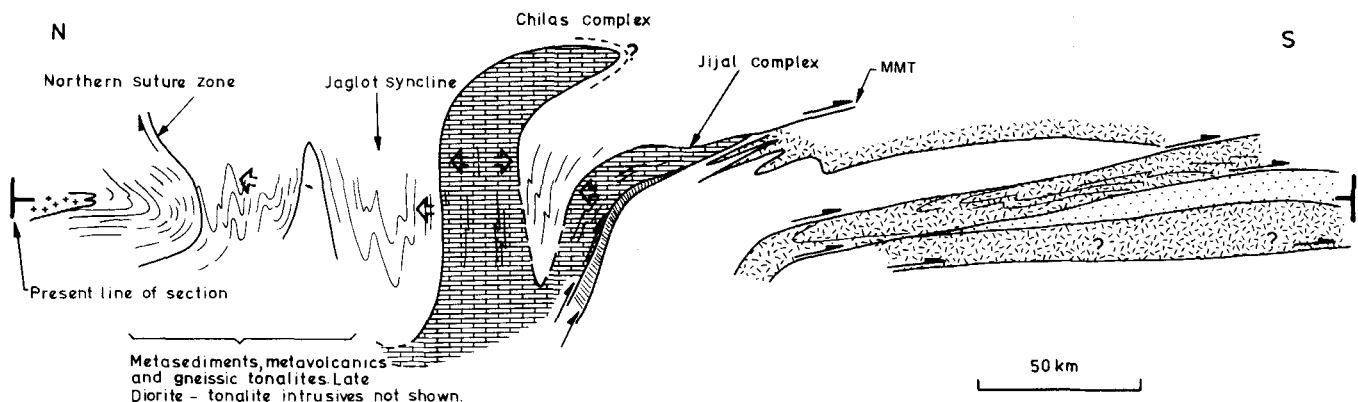


Fig. 2 Simplified tectonic cross-section of the region. For location and caption see Fig. 1. The dip and position of the major thrusts at depth is uncertain. Evidence for flat-lying thrusts beneath the Indian plate is given in ref. 13. Vertical scale = horizontal scale.

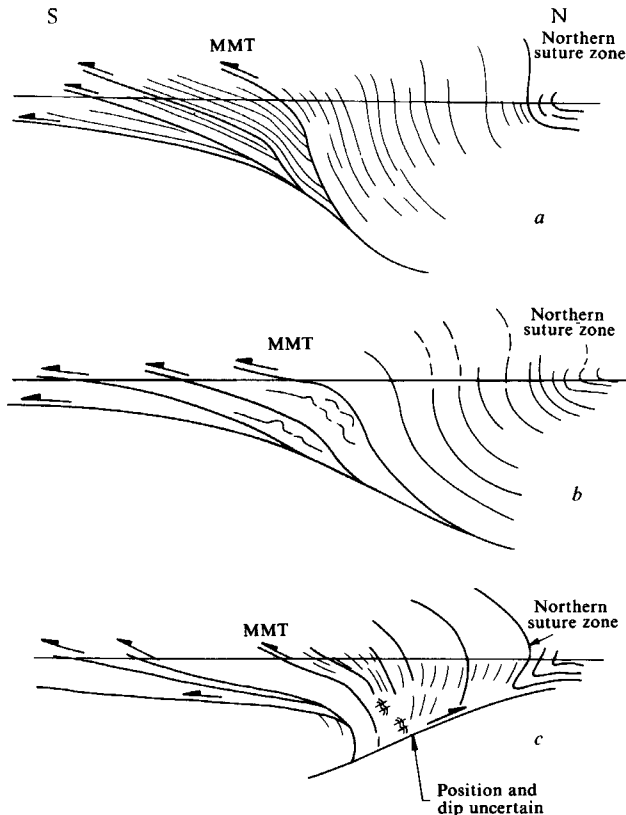


Fig. 3 Model section to show how Kohistan sequence achieved its steep dip: *a*, due to climb and fanning of thrusts to the south; *b*, due to folding above a decoupling horizon⁹; *c*, due to back thrusting. The structure of the Kohistan complex is simplified to show approximate dips of foliation and fold axial planes. Vertical scale = horizontal scale.

granite which has a Rb/Sr whole rock isochron age of 516 ± 16 Myr (ref. 8). This granite has a contact aureole against pelitic sediments of presumed late Precambrian age. These basement rocks are thrust southwards over Mesozoic sediments near Abbottabad and together with the Mesozoic sediments they are further thrust over mid-Oligocene to mid-Miocene red sandstones and shales (the Murree Formation). All these rocks are folded into a major steeply-plunging tight anticline (Fig. 1) which is a first phase structure in the Miocene and at least a third phase structure in the Mesozoic sediments and their gneissic basement. To the south there are even more thrusts in sediments ranging in age from Palaeocene to Cambrian, flooring in a blind thrust in the sediments above the basement.

Discussion

The MMT clearly separates crustal blocks of two different ages, origin and lithology^{1,2}. The gneisses to the south form part of the late Precambrian-early Palaeozoic segment of the Indian plate, while to the north no old basement has been found, but rather deformed sediments and igneous rocks of Mesozoic-Cenozoic age. The Mesozoic-Cenozoic Kohistan sequence was considered to represent an island arc^{1,2}, but our studies show this to have a more complicated structural history and a compound origin.

The lower crust is made up of the basic-ultrabasic Chilas complex which is >300 km long and >8 km thick, and is thus

one of the largest intrusions in the world. Above this there are mainly pillow lavas and greywackes intruded by gabbros, diorites and tonalites. The early isoclines are developed in the lower crustal rocks of the southern Kohistan sequence associated with granulite facies metamorphism. These structures are folded by one of the largest synclines known, which folds the whole upper crust of Kohistan. Its synclinal axis changes from near horizontal in the west to steeply NNE near the Indus River parallel to the Himalayan overthrust direction. This major syncline and related structures seem to be related to the MMT, though the eastern continuation of the MMT is difficult to trace around the Nanga Parbat syntaxis which has the form of a late fold with late faults along its western margin. In Kohistan the MMT marks a major suture bringing granulites against blueschists and amphibolites. The immediately underlying Indian plate contains a 10–20 km wide, gently dipping shear zone with Mesozoic cover tightly infolded into basement gneiss. To the south there are more discrete thrust planes, the floor thrust of which climbs southwards from basement to progressively higher levels of sediment. These structures propagated southwards: the earlier thrusts and folds in the north are refolded by structures developed in the south. The Kohistan sequence north of the MMT has been backfolded to become vertical to southerly dipping west of the Nanga Parbat syntaxis. The origin of this backfold is uncertain: one method of modelling would be to produce the above-described major thrusts in the form of an imbricate fan analogous to that in accretionary wedges⁹ (see Fig. 3*a*). Alternatively, the more northern thrusts could be folded into a steep orientation above a low-level decoupling horizon analogous to that suggested for the Swiss Alps (Fig. 3*b*)¹⁰. In both these models the depth to the decoupling zone would be at least 100 km below the present erosion surface (Fig. 3*a, b*). In the Ladakh area, however, 200 km along the strike to the east of Kohistan, there are southward-dipping faults overthrusting to the north¹¹, and similar thrusts occur along the Indus suture in the central Himalaya¹². Possibly the backfold is related to uplift on these faults (Fig. 3*c*).

Conclusion

One important conclusion from our study is that the predominant diorite-tonalite igneous activity in northern Kohistan is late to post-tectonic; some of it certainly postdates the Jaglot syncline and may postdate the backfolding. This conclusion is supported by a preliminary study of the Skardu area, east of the Nanga Parbat, where most granites, tonalites and gabbros post-date the intensely deformed metasediments and volcanics. This throws doubt on the contribution of this igneous activity to the development of the island arc. The problem of the origin of these diorites and tonalites may be solved by isotope and trace element geochemistry, at present underway at Cambridge, Leeds and Leicester.

The other exciting feature of the Kohistan area is the size of the structures involved. The Kohistan complex is now up to 150 km thick though this is largely the result of two phases of folding; little evidence has been found for thrust repetition as originally suggested by Bard *et al.*². The Jaglot synform is one of the largest reported structures in the world and involves deformation of the whole crust with decoupling at considerable depth (>100 km) in the mantle. Similar decoupling depths are indicated by the different models to explain the backfolding (Fig. 3).

We thank NERC for research grant GR3/4242, and R. A. K. Tahirkheli for logistical support.

Received 10 August; accepted 5 November 1981.

1. Tahirkheli, R. A. K., Mattauer, M., Proust, F. & Tapponnier, P. in *Geodynamics of Pakistan* (eds Farah, A. & DeJong, K. A.) 125–130 (Geological Survey of Pakistan, 1979).
2. Bard, J. P., Maluski, H., Matte, P. H. & Proust, F. in *Geol. Bull. Univ. Peshawar Spec. Iss.* **13**, 87–94 (1980).
3. Tapponnier, P., Mattauer, M., Proust, F. & Cassaigne, C. *Earth planet. Sci. Lett.* **52**, 355–371 (1981).
4. Tahirkheli, R. A. K. & Jan, M. Q. *Geol. Bull. Univ. Peshawar Spec. Iss.* **11** (1979).

5. Khan, J. *Geol. Bull. Univ. Peshawar Spec. Iss.* **11**, 153–162 (1979).
6. Jan, M. Q. & Howie, R. A. *J. Petrol.* **22**, 85–126 (1981).
7. Shams, F. A., Jones, G. C. & Kempe, D. R. C. *Miner. Mag.* **43**, 941–942 (1980).
8. Le Fort, P., Debon, F. & Sonet, J. *Geol. Bull. Univ. Peshawar Spec. Iss.* **13**, 51–61 (1980).
9. Moore, J. C. & Karig, D. E. *Bull. geol. Soc. Am.* **87**, 1259–1268 (1976).
10. Hsu, K. J. *Am. J. Sci.* **279**, 353–366 (1979).
11. Frank, W., Gansser, A. & Trommsdorff, V. *Schweiz. miner. petrogr. Mitt.* **57**, 89–113 (1977).
12. Shackleton, R. M. *J. struct. Geol.* **3**, 97–105 (1981).
13. Seeber, L. & Armbruster, J. in *Geodynamics of Pakistan* (eds Farah, A. & DeJong, K. A.) 131–142 (Geological Survey of Pakistan, 1979).

Cellular origin and role of mink cell focus-forming viruses in murine thymic lymphomas

Sisir K. Chattopadhyay^{*†||}, Miles W. Cloyd[‡], David L. Linemeyer[§], Marilyn R. Lander^{*†}, Elaine Rands^{*} & Douglas R. Lowy^{*}

^{*} Dermatology Branch, [†] Pediatric Oncology Branch and [§] Laboratory of Tumour Virus Genetics, National Cancer Institute, Bethesda, Maryland 20205, USA

[‡] Rocky Mountain Laboratory, National Institute of Allergy and Infectious Diseases, Hamilton, Montana 59840, USA

Embryo DNA of AKR mice contains several copies of intact mink cell focus-forming (MCF)-like env gene sequences with a proviral structure. In spontaneous thymic tumours, these MCF-like env sequences have regularly recombined with a specific region of the 3' end of ecotropic env. This specific recombination is probably a critical event in the pathogenesis of spontaneous AKR tumours.

MURINE leukaemia viruses (MuLVs) are endogenous to laboratory mice¹⁻³. The MuLVs of these mice have been divided into three related classes on the basis of their host range in tissue culture, which is determined by the viral envelope glycoprotein (gp70) encoded by the viral gene *env*⁴. These classes are: ecotropic, which infect only murine cells; xenotropic, which replicate primarily in heterologous cells; and mink cytopathic focus-forming (MCF) viruses, which grow in both murine and heterologous cells⁵⁻⁸. Biochemical studies of viral nucleic acids have suggested that MCF viruses are recombinants between ecotropic MuLV and endogenous *env* sequences related to xenotropic viruses^{9,10}.

Among laboratory mice, the AKR strain¹¹ has been intensively studied because of its high incidence of MuLV-associated spontaneous T-cell leukaemias, which originate in the thymus. Some steps leading to the development of this thymic disease have been determined, but many aspects of this complex process remain to be elucidated¹². Each AKR mouse cell contains in its chromosomal DNA one or more complete copies of an ecotropic MuLV genome (proviral DNA)¹³⁻¹⁶. This viral genome is usually spontaneously activated shortly before birth, leading to high levels of circulating virus throughout the animal's life¹⁷. This ecotropic virus expression is apparently necessary but not sufficient for tumour development, which occurs only after the animal is at least 6 months old¹⁸. MCF viruses have been isolated primarily from preleukaemic thymuses and from tumour tissue^{5,6}, suggesting that MCF viruses may represent a more proximate cause of the tumour.

The origin and role of MCF viruses in tumorigenesis, which we have studied here, are still incompletely understood. The endogenous non-ecotropic viral sequences which presumably participate in the formation of MCF viruses have not been identified in normal cells, but it is known that AKR and other mice contain multiple copies^{2,3} of sequences homologous to part of the complete AKR ecotropic viral genome. In addition, only certain MCF viruses are oncogenic¹⁹. It therefore appears that virological factors in addition to the expanded *in vitro* host range have a significant role in tumorigenesis.

In this study, we have used restriction endonucleases and a viral probe specific for MCF and xenotropic *env* sequences to map the viral sequences in high molecular weight DNA from AKR mouse embryos and from spontaneous and MCF virus-induced tumours. These maps have been compared with those recently developed for the DNAs of MCF viruses²⁰, where consistent differences were noted between DNAs of pathogenic and non-pathogenic MCF viruses. We have found: (1) AKR embryo DNA contains ~10 copies of apparently complete *env*

sequences closely related to those found in MCF viral genomes; (2) these endogenous MCF-like *env* sequences resemble the *env* of the non-pathogenic MCF viruses more closely than they do those of pathogenic MCF viruses; (3) at least some copies of these *env* sequences have the structure of endogenous proviruses; (4) spontaneous AKR tumours are clonal in origin and contain recombinant viral DNA with the same structure as that found in pathogenic MCF viral DNAs. The results suggest that specific recombinational events between the endogenous MCF-like sequences and ecotropic viral sequences are a prerequisite for the establishment of AKR tumours.

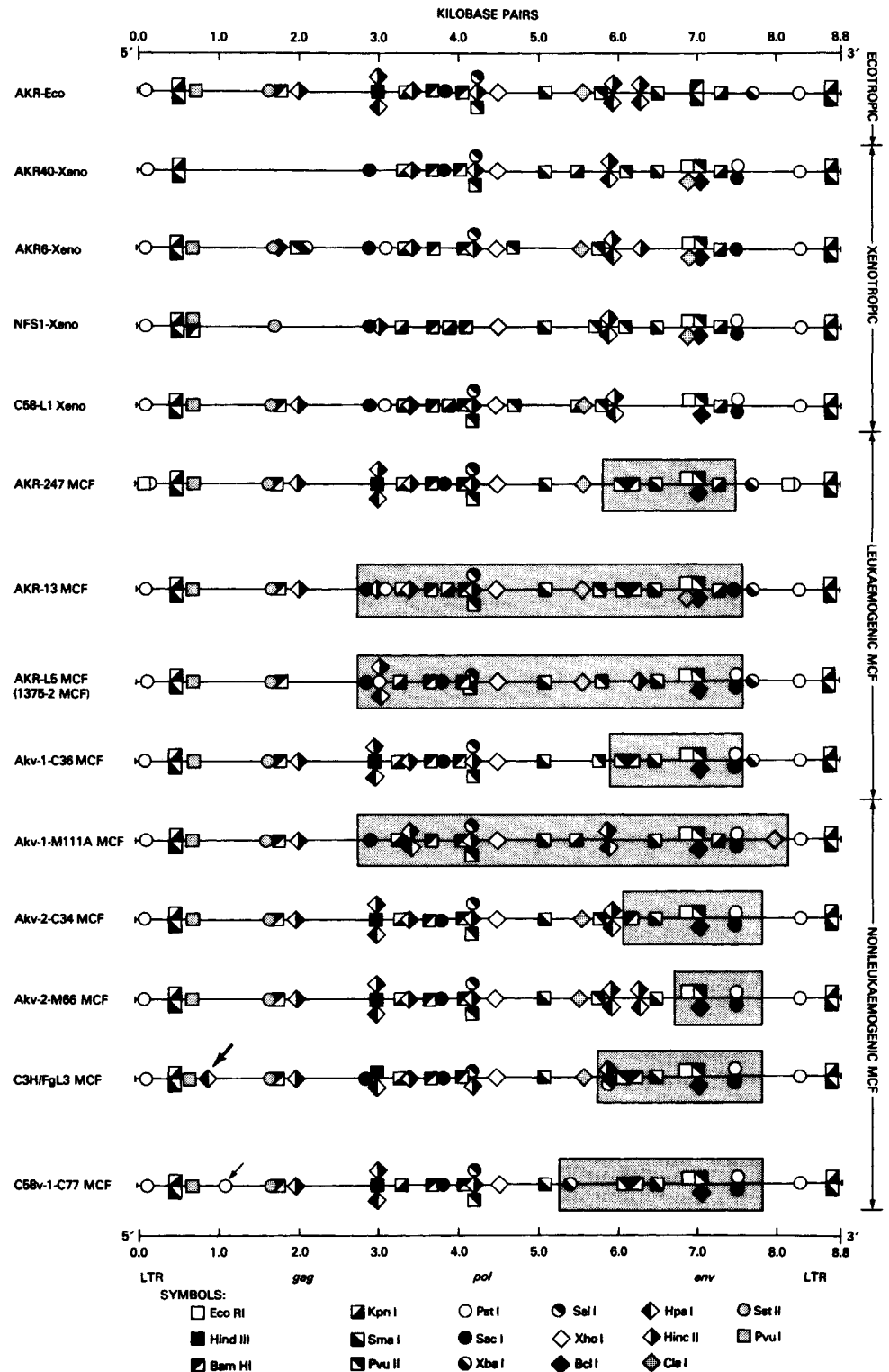
Physical maps of ecotropic, xenotropic and MCF viral DNAs

Before analysing the AKR high molecular weight DNA, we will summarize the results of physical mapping studies of unintegrated DNAs of ecotropic, xenotropic and MCF viruses from AKR and other strains so that these maps can be compared with those generated here for endogenous non-ecotropic viral sequences in the high molecular weight DNA (Fig. 1)^{20,21}. The viral DNAs, which are 8.8 kilobases (kb), contain three genes: *gag* (spanning 1.0–2.7 kb; map locations are given by their distance from the left (5') end of the viral DNA), which specifies the virion core proteins, *pol* (2.7–5.8 kb), which encodes the reverse transcriptase, and *env* (5.8–8.2 kb), which codes for the virion envelope proteins gp70 and p15E (see Fig. 1) (refs 22–26 and C. Van Beveran and I. M. Verma, personal communication). The p15E sequences, which begin at ~7.5 kb, are located to the right of those which encode gp70 (refs 24–27 and C. Van Beveran and I. M. Verma, personal communication). The viral DNAs are flanked at each end by 0.6 kb of repeat sequences called the long terminal repeat (LTR).

As predicted from earlier studies of viral RNA oligonucleotides²⁸, each MCF viral DNA was unique, each was apparently composed of ecotropic and non-ecotropic sequences, and each contained non-ecotropic *env* sequences which shared restriction endonuclease sites with xenotropic *env* sequences (such as *EcoRI* at 6.9 and *BclI* at 7.0 kb, Fig. 1). However, some sites (such as *BamHI* at 6.2 kb) were unique to MCF viruses, suggesting that the non-ecotropic sequences of MCF viruses might not be derived from proviruses of known xenotropic viruses. In addition, a consistent difference between the viral DNAs of non-pathogenic and pathogenic MCF viruses was that pathogenic MCF viral DNAs contained an *XbaI* site in p15E at 7.7 kb which was apparently derived from the ecotropic viral parent, while non-pathogenic MCF viruses did not have this site. Finally, in certain MCF viral DNAs, the span of the inferred segment of non-ecotropic sequences was much greater than only part of the gp70 coding region (Fig. 1). These results suggested

|| To whom reprint requests should be addressed at: Building 41, Room A114, National Institutes of Health, Bethesda, Maryland 20205, USA.

Fig. 1 Comparison of linear restriction endonuclease maps of representative ecotropic, xenotropic and MCF murine leukaemia viral DNAs. The viral DNAs are grouped according to the viral host range and within MCF viruses according to their disease-inducing capability. Each symbol corresponds to an enzyme as indicated in the figure. AKR-eco (ecotropic virus isolated from an AKR mouse) is virtually identical to ecotropic viruses of many inbred strains of mice (the strains and isolates analysed are: *Akv-1*, two isolates; *Akv-2*, two isolates; C58, three isolates; C3H/Fg, three isolates; PL, three isolates; F, one isolate; BALB/c, one isolate) and of *Mus musculus molossinus*^{20,21}. Nomenclature and origin of various viruses have been described previously^{20,21}. Each MCF DNA has been assumed to contain a continuous insert of non-ecotropic sequences, which may represent oversimplification in some instances. In general, the length of this insert (depicted by the wide bar) represents a minimum estimate, based on the presence of xenotropic or MCF-specific sites and/or the loss of sites common to ecotropic and xenotropic DNAs. The *HpaI* site at 0.8 kb in C3H/FgL3 MCF and the *PstI* site at 1.0 kb in C58v-1-C77 MCF (shown by arrows) are not found in ecotropic MuLVs, but have not been considered in the estimated segment of non-ecotropic sequences. These maps were derived from the unintegrated proviral DNAs as described in detail elsewhere^{20,21,50}. Briefly, cells were infected with MuLV from chronically infected cells and DNA was extracted from them by the Hirt⁵¹ procedure 18 h after infection. DNA from the Hirt supernatant was used as the starting material for restriction endonuclease digestion. Unintegrated proviral DNAs were digested with various enzymes, electrophoresed, transferred to nitrocellulose filters⁵² and hybridized with ³²P-cDNA synthesized either from 70S viral RNA or by nick-translation of molecularly cloned viral DNAs. The cleavage products of each enzyme were ordered and oriented for each viral DNA by hybridization of digested DNAs with a 5' probe (the 4.6-kb *SmaI* fragment spanning 0.5–5.1 kb of AKR ecotropic MuLV DNA), a 3' probe (the 1.7-kb *SmaI* fragment spanning 7.0–8.7 kb of the ecotropic DNA) derived from molecular clone AKR MuLV 623 DNA⁵⁰ or a full xenotropic or MCF probe after single and/or double digestion of various Hirt DNAs with restriction endonucleases.



that the gp70 (and p15E) of at least some MCF viruses might be derived entirely from non-ecotropic sequences and that the *env* sequences which encode the expanded *in vitro* host range of MCF viruses might be found in high molecular weight DNA of normal cells.

Isolation of a DNA fragment specific for MCF and xenotropic *env*

To probe the high molecular weight DNA for the endogenous sequences which might give rise to the non-ecotropic portion of MCF viruses, we cloned in pBR322 a viral DNA *env* fragment which hybridized to MCF and xenotropic viral DNAs but not to

ecotropic viral DNAs. This viral probe, designated BE, is composed of the 0.7 kb between a *Bam*HI and *Eco*RI site in *env* of the Friend spleen focus-forming virus (SFFV) DNA²⁹, a helper-dependent MuLV whose *env* is closely related to those of MCF viruses³⁰. This region in SFFV corresponds exactly to the segment in some MCF viral DNAs between the *Bam*HI site at 6.2 kb and the *Eco*RI site at 6.9 kb (see diagram in Fig. 2). The BE fragment was characterized by radioactively labelling it and incubating it with Hirt supernatants containing ecotropic, xenotropic or MCF viral DNAs: the probe hybridized strongly to MCF DNAs and moderately to xenotropic DNAs, but not to ecotropic DNAs (Fig. 2). Similar results were obtained with the analogous *Bam*HI-*Eco*RI fragment derived from cloned MCF 247 DNA (data not shown).

Embryo DNA contains MCF-like *env* sequences

We then used the BE fragment to probe AKR embryo DNA for the presence of endogenous sequences from which the non-ecotropic *env* sequences of MCF viruses should have been derived (Fig. 3b, c). Hybridization of the BE probe to *Bam*HI-digested embryo DNA was used to estimate the number of copies of cross-reacting sequences in this DNA, as the BE probe should hybridize to cell-virus DNA junction fragments in high molecular weight DNA digested with this enzyme. *Bam*HI digestion (Fig. 3b) yielded ~12–15 fragments, indicating that there are multiple copies in embryo DNA of sequences which cross-hybridize with the BE probe.

We had previously found that *env* of many non-pathogenic MCF viral DNAs had a *Bam*HI site at 6.2 kb, all had an *Eco*RI site at 6.9 kb and a *Bcl*I site at 7.0 kb, some had a *Kpn*I site at 7.3 kb and all had a *Pst*I site at 7.5 kb, a *Kpn*I site at 7.3 kb, a *Pst*I site at 7.5 kb and a *Kpn*I site at 8.7 kb (in the LTR) (see Figs 1, 3). Several copies having these sites were also present in the embryo DNA.

For example, *Eco*RI–*Bcl*I digestion (Fig. 3b) shortened each *Bcl*I fragment by 0.1 kb, which indicated that each fragment contained an *Eco*RI site 0.1 kb away from a *Bcl*I site, as was also the case in *env* of every xenotropic and MCF viral DNA. *Bam*HI–*Eco*RI digestion (Fig. 3c) gave a more simplified pattern than did *Bam*HI digestion; the intense 0.7-kb fragment corresponded to the *Bam*HI–*Eco*RI fragment spanning 6.2–6.9 kb in many MCF viral DNAs, suggesting the presence of several copies of this fragment in embryo DNA. Endogenous fragments which contained the *Kpn*I site at 7.3 kb should yield a 1.1-kb *Bam*HI–*Kpn*I fragment (6.2–7.3 kb), while those lacking this *Kpn*I site should yield a 2.5-kb *Bam*HI–*Kpn*I fragment (6.2–8.7 kb); these are the two prominent fragments of this double digestion (Fig. 3c). The 7.5-, 2.9- and 2.4-kb *Pst*I fragments were apparently all shortened by 0.6 kb when cleaved with *Eco*RI (Fig. 3b), which suggests that the right end of these *Pst*I fragments all terminate at 7.5 kb.

It was also important to determine if the *Bam*HI fragments contained an *Xba*I site corresponding to that found in p15E at 7.7 kb in pathogenic MCF viruses. Hybridization of the BE probe to *Bam*HI–*Xba*I-digested AKR embryo DNA (Fig. 3c) did not yield a 1.5-kb fragment which would correspond to the *Bam*HI–*Xba*I fragment found in pathogenic MCF viruses such as MCF247. The *Xba*I site at 7.7 kb noted in ecotropic viruses and in pathogenic MCF viruses was therefore not covalently linked to the 6.2-kb *Bam*HI site in the embryo DNA. Taken together, the results described above demonstrated that AKR embryo DNA contained multiple (we estimate ~10) copies of apparently complete *env* sequences with the same restriction sites mapped in non-pathogenic MCF viral DNAs, and at least some of these *env* sequences were linked on their right to an LTR. It is probable that several of these copies can participate in the formation of MCF viruses, as their heterogeneity with respect to the *Bam*HI site at 6.2 kb is correlated with heterogeneity at this site in MCF viruses (Fig. 1).

Endogenous MCF-like *env* sequences are part of proviral DNAs

We have also found evidence that at least some of these MCF-related *env* sequences are part of proviral DNAs, as they are also linked on the left to LTR, *gag* and *pol* sequences. As seen in Fig. 2 and Fig. 3b, the BE probe hybridized to a prominent 7.5-kb *Pst*I fragment, which is the expected length of a proviral fragment spanning sequences from a *Pst*I site near the left end of the LTR to a *Pst*I site at 7.5 kb (both of these *Pst*I sites have been found in xenotropic and MCF viral DNAs, see Figs 1, 3). Most of the copies that make up this *Pst*I fragment do extend to 7.5 kb, because, as noted previously, *Eco*RI digestion apparently shortens this fragment by 0.6 kb. In addition, this 7.5-kb *Pst*I fragment hybridizes to probes specific for the 5' end (5' probe), middle portion (M probe) and LTR (L probe) of viral

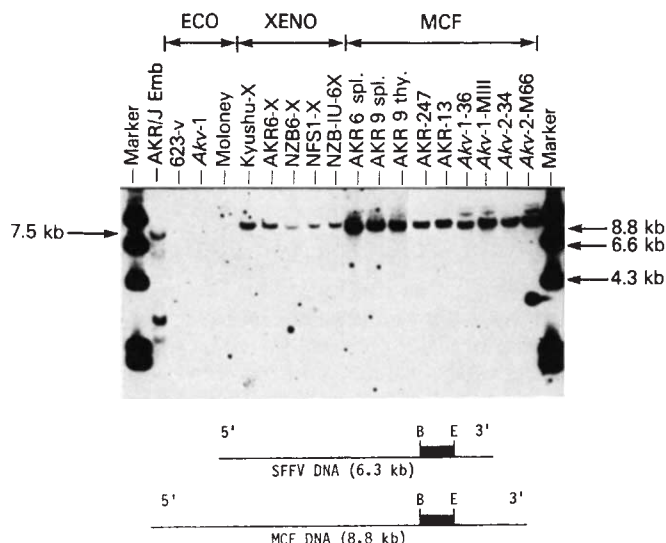


Fig. 2 Southern blot showing selective hybridization of the BE probe to MCF and xenotropic viral DNAs, but not to ecotropic viral DNAs. The BE probe, derived from a 0.7-kb *Bam*HI–*Eco*RI fragment of Friend SFFV DNA²⁹, was molecularly cloned (by D.L.L.) in pBR322 via the *Bam*HI and *Eco*RI sites. In the upper diagram, the location of the BE probe sequences in SFFV DNA is represented by the heavy line between B (= *Bam*HI) and E (= *Eco*RI). Also shown are the sequences in the BE probe which correspond to the 6.2–6.9-kb region in MCF viruses in *env*, as described in the text. Unintegrated ecotropic (eco), xenotropic (xeno) and MCF viral DNAs were isolated by the Hirt procedure 16–18 h after infection of permissive cells as described in Fig. 1 legend. Hirt supernatant DNAs (5–7 µg each for ecotropic viruses, 15–20 µg of each for xenotropic and MCF viruses) were electrophoresed in horizontal 0.7% agarose gels (35 V, 16 h), transferred to nitrocellulose filters and hybridized with ³²P nick-translated 0.7-kb BE probe. The amount of Hirt DNA added for each virus gave an 8.8-kb band of similar intensity when hybridized to its homologous probe (data not shown). *Pst*I-digested high molecular weight AKR/J embryo DNA was electrophoresed in the lane marked AKR/J Emb. 623-v indicates the 623 ecotropic MuLV isolated from NIH3T3 cells transfected with AKR MuLV 623 DNA. The origin of the other viruses has been described previously. Marker was *Hind*III-digested ³²P-labelled λ DNA. The numbers beside the panels indicate the length of fragments in kb.

DNAs (Fig. 3a). (The 8.2-kb *Pst*I fragment which hybridized to each probe except the BE probe represents the endogenous ecotropic proviral DNA cleaved at the *Pst*I site in each LTR¹⁵.) Several *pol*–*env* fragments which corresponded to fragments in MCF viral DNAs in Fig. 1 whose non-ecotropic sequences were inferred to include the *pol* region were also noted in embryo DNA. For example (Fig. 3b), the 5.3-kb *Kpn*I fragment corresponded to a fragment in L5 MCF (3.4–8.7 kb) and the 1.8-kb *Kpn*I fragment to a fragment in M111A MCF (5.5–7.3 kb).

The above results therefore indicate that at least some of the endogenous sequences which hybridize to the BE probe are organized as proviral DNAs, as already shown for mouse ecotropic^{15,16} and avian leukosis³¹ viral genomes in embryo DNA. We have also found multiple copies of similar MCF-like sequences in other mouse strains, including BALB/c and NIH Swiss, and have molecularly cloned from high molecular weight DNA of BALB/c and NIH cells *Eco*RI fragments which hybridize strongly to the BE probe: each clone contains cell sequences linked to 6.9 kb of proviral DNA truncated at the (6.9 kb) *Eco*RI site (unpublished data).

Tryptic analysis of viral gp70s has suggested that this protein in MCF viruses is part ecotropic and part xenotropic, which has led to the hypothesis that the MCF gp70 is invariably a hybrid gene product³². However, the finding that the endogenous MCF-like *env* sequences closely resemble those in MCF viruses supports the alternative hypothesis^{20,33} that the MCF phenotype assayed in culture may be encoded completely by the endogenous non-ecotropic *env* sequences. The development of MCF-specific gp70 assays^{30,34–36} is also consistent with this hypothesis. Further, the MCF gp70 tryptic peptides which resemble ecotropic peptides, which have been localized to the amino terminus of this protein³⁷, are encoded in the BE probe, and the *env* sequences in the truncated BALB/c and NIH Swiss

clones mentioned above resemble the *env* of MCF247 in that they contain the same *PvuII*, *BamHI*, *SmaI* and *EcoRI* sites between 6.1 and 6.9 kb (unpublished data). These data suggest that ecotropic-like gp70 peptides of the MCF viruses may be derived from the endogenous MCF-like *env* sequences. However, a direct answer must await biological studies of these endogenous sequences after they have been molecularly cloned intact.

AKR tumours are clonal and contain new MCF DNAs

Having characterized the endogenous sequences which hybridized to the BE probe, we examined the viral sequences in high molecular weight DNA from 19 spontaneous AKR tumours and from 11 tumours induced (accelerated) in AKR mice with MCF247 virus. Other studies with unfractionated probes have found new viral DNA fragments in tumour DNA³⁸⁻⁴⁰. The DNAs from MCF virus-induced tumours were used principally as controls for the spontaneous tumours, as it is known that the inoculated MCF virus can be recovered intact from these tumours³⁸.

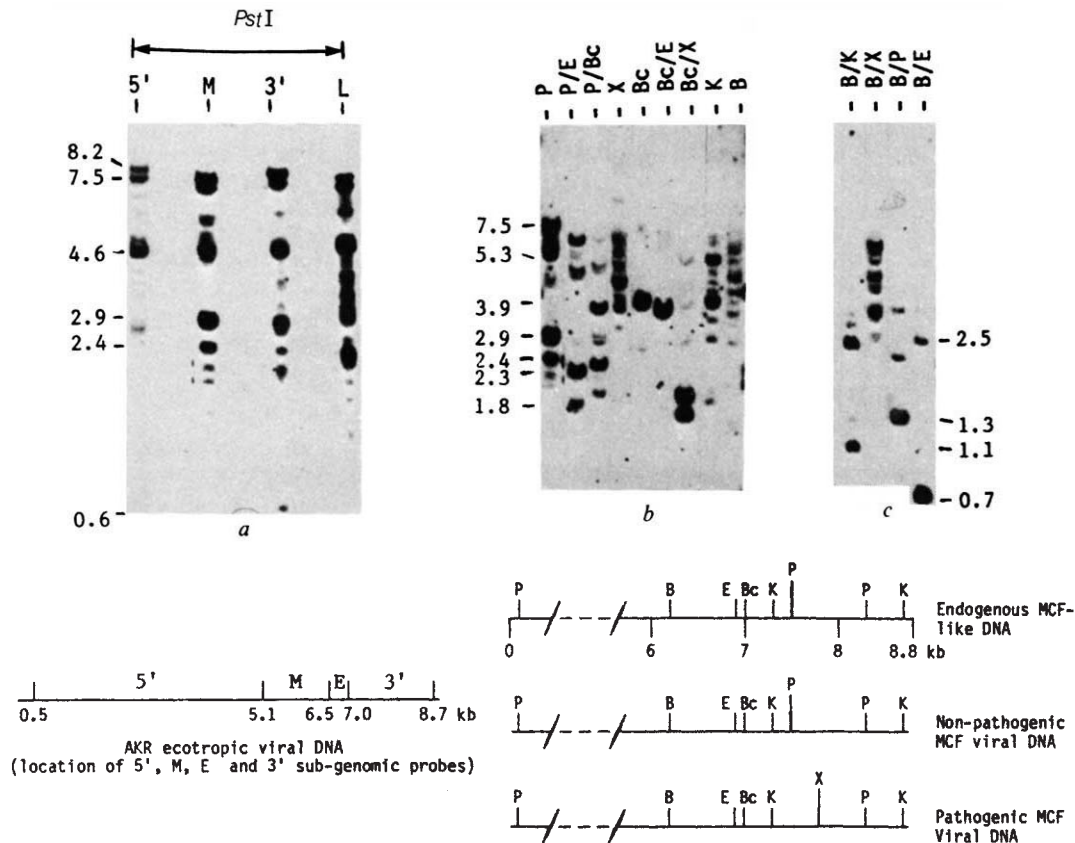
The tumour DNAs were digested with *BclI*-*XbaI* and hybridized to three viral probes: the 5' probe, to detect new MuLV sequences of any tropism; an ecotropic specific probe (the E probe shown below Fig. 3a) to detect viral DNAs with ecotropic gp70 sequences; and the BE probe, to detect viral DNAs with MCF gp70 sequences. As all three probes are composed of sequences located to the left of 7.0 kb, while the only *BclI* and *XbaI* sites in AKR MCF or ecotropic viral DNAs were located to the right of 7.0 kb, this digestion should reveal new left-hand cell-virus junction fragments with each probe.

Representative results are presented in Fig. 4. When thymic tumour DNA digested with *XbaI*/*BclI* was hybridized with the 5' probe, novel fragments were detected in all MCF virus-induced tumours (Fig. 4a, tumours I 01 and I 02) and in eight of

the spontaneous tumours. In Fig. 4a, one spontaneous tumour contained a novel fragment (S 02, arrow), while the other spontaneous tumour did not (S 01). Others^{39,40} have also found new viral fragments in only some spontaneous AKR tumours, probably because of the high background of multiple endogenous fragments. The tumours induced with MCF247 virus contained several new copies of viral DNA, as denoted by haziness and new distinct bands of 7-9 kb. By contrast, the spontaneous tumours usually contained only a single novel fragment. For each MCF-induced tumour DNA, most of the new viral fragments hybridized to the BE probe (Fig. 4b), but other fragments hybridized to the ecotropic probe (Fig. 4d), indicating that the tumours were infected with both MCF and ecotropic viruses. By contrast, the new fragment detected with the 5' probe in spontaneous tumours also hybridized with the BE probe (Fig. 4b, tumour S 02, arrow), but no new ecotropic fragments were detected in spontaneous tumours (Fig. 4d), despite the small numbers of endogenous ecotropic fragments. These results suggest that ecotropic viruses as such do not usually infect the cells before they become a tumour. (Since submission of this article, Yoshimura and Breda⁴¹ have reported similar observations.) Thymic lymphocytes have been reported to be resistant to infection by ecotropic virus, and our failure to find additional ecotropic viral genomes in the spontaneous tumours is consistent with this observation^{42,43}.

We also determined that the AKR tumours were, as previously suggested^{38,39}, probably of clonal origin, as has already been demonstrated for avian leukosis virus-induced bursal tumours in chickens^{44,45}. Each new fragment in the DNA of a given thymoma was found to be characteristic of all tumour tissue from a given animal, whether the tumour arose spontaneously or was induced by MCF247 (Fig. 4c, arrow). Note also that uninvolved intestinal tissue from leukaemic animals did not contain these new fragments. As the metastatic tumours contained the same new viral fragments as the primary

Fig. 3 Blot analysis of nonecotropic sequences of AKR/J embryo DNA. AKR/J embryo DNA was digested with various enzymes as indicated, electrophoresed in 0.5% agarose gels (35 V, 24 h), UV irradiated for 10 min, transferred to nitrocellulose filters and hybridized with various viral probes as described below. Fragment lengths are given in kb. In *a*, each lane contained *PstI*-digested embryo DNA. The lanes labelled 5', M and 3' were hybridized to probes derived from cloned *SmaI* fragments of AKR ecotropic viral DNA as shown in the diagram below *a*. The 5' probe spans 0.5-5.1 kb, the M probe spans 5.1-6.5 kb and the 3' probe spans 7.0-8.7 kb. The lane in *a* labelled L was hybridized to a probe composed of the LTR sequences of Harvey murine sarcoma viral DNA (provided by E. H. Chang, NCI); this probe also hybridizes to the LTRs of the viruses isolated from AKR mice. In *b* and *c*, the embryo DNA was hybridized to the BE probe, described in Fig. 2 legend. In these two panels and in the diagram below them, K = *KpnI*, P = *PstI*, B = *BamHI*, X = *XbaI*, E = *EcoRI*, Bc = *BclI*. The deduced restriction endonuclease map of the non-ecotropic viral sequences in the embryo DNA has been shown in the map labelled 'endogenous MCF-like viral DNA'. These sequences have been compared with those found in non-pathogenic MCF viral DNAs. Viral sequences located rightward from the *BamHI* site at 6.2 kb have been emphasized here. However, at least some of these endogenous MCF-like DNAs are organized as proviral DNAs (LTR, *gag*, *pol*, *env*, LTR), as noted in the text.



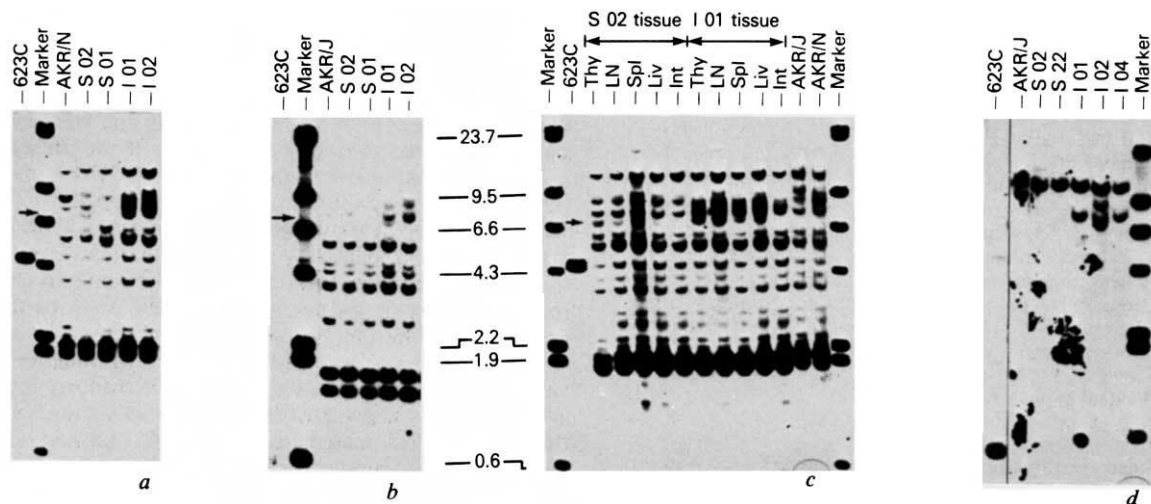


Fig. 4 AKR tumours are clonal in origin, contain new MCF viral DNA, but not new ecotropic viral DNA. *a* Shows that new viral DNA fragments are detected in some spontaneous tumours and all MCF virus-induced tumours; *b* shows that in spontaneous tumours the new fragments hybridize with the BE probe; *c* shows the clonal nature of the tumours; *d* shows that spontaneous tumours do not contain new ecotropic viral DNAs, but that MCF virus-induced tumours do. High molecular weight DNAs isolated from various spontaneous tumours (S) or MCF-induced tumours (I) were digested with *Bcl*I-*Xba*I, electrophoresed in 0.5% agarose gel (35 V, 24 h), UV irradiated for 10 min, transferred to nitrocellulose filters and hybridized with 32 P nick-translated viral probes: the 5' probe (described in Fig. 3 legend) in *a* and *c*, the BE probe (described in Fig. 2 legend) in *b*, or an ecotropic-specific probe spanning 6.6–7.0 kb in AKR ecotropic viral DNA (labelled E in the diagram below Fig. 3a) in *d*. *Hind*III-digested 32 P-labelled λ DNA was used as marker and the number indicates length of the DNA in kb. In *a*, *b* and *d*, DNAs from thymic tumours were used. AKR/J and AKR/N indicates DNAs from embryos of these strains respectively. In *c*, Thy = thymus, LN = lymph nodes, Spl = spleen, Liv = liver, Int = intestine. Thymus, lymph nodes, spleen and liver contained tumours, while intestine did not. In each panel, the lane marked 623C represents *Sma*I-digested clone 623 AKR ecotropic viral DNA as a specificity control. In *d*, the AKR/J embryo DNA contains one more ecotropic fragment than does the DNA from the spontaneous tumours. This additional copy is present only in some AKR/J embryos (M. Martin, personal communication).

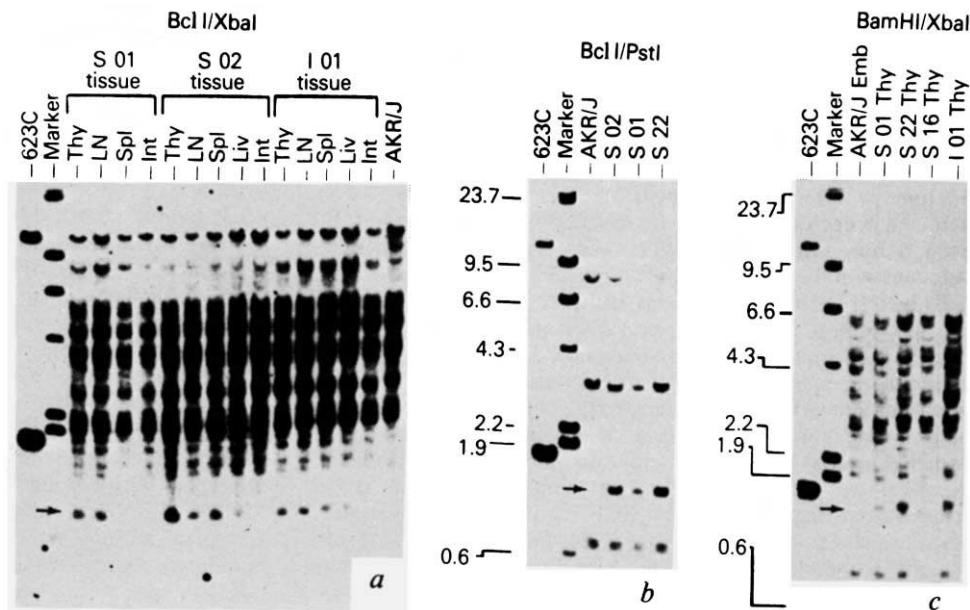


Fig. 5 Blot analysis of tumour DNAs showing that they contain recombinant viral sequences. S = spontaneous; I = induced; Thy = thymus; Spl = spleen; LN = lymph node; Liv = liver; Int = intestine. Thymic tissue was used for tumours in *b*. In each panel, the lane marked 623C contains *Sma*I-digested clone 623 viral DNA as a specificity control. High molecular weight DNAs were digested with various combinations of restriction endonucleases as indicated at the top of each panel, electrophoresed in 0.5% agarose gel (35 V, 24 h), UV irradiated, transferred to nitrocellulose filters and hybridized with a 32 P-labelled nick-translated 3' probe (described in Fig. 3). In the diagram, the restriction enzyme sites of the recombinant viral *env*-LTR are shown on the map labelled 'recombinant DNA in tumours' (the presence of the *Kpn*I sites has not been determined experimentally).

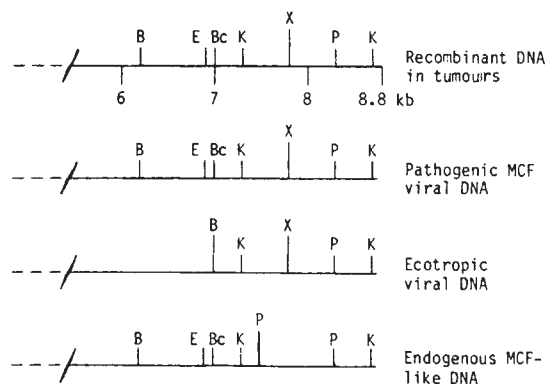


Table 1 MCF virus in AKR thymus

	Infectious centres per 10 ⁷ cells
AKR247 MCF-induced thymomas (19)	10,000–85,000
AKR spontaneous thymomas (7)	Trace, 1,000
Pre-leukaemic AKR (<6 months old) (9)	0*

Mink lung cells (CCL64) were plated at 2×10^5 cells per 60-mm dish in Dulbecco's modified minimal essential medium with 10% fetal calf serum. The following day, serial dilutions of thymus or spleen cells were plated as infectious centres onto the mink cultures in medium containing $16 \mu\text{g ml}^{-1}$ polybrene. The medium was changed 24 h later and the cultures were scored for characteristic MCF cytopathic foci (mink CPE) after confluence (days 5–7). The numbers in parentheses indicate the number of individual animals tested.

* Thymus and spleen.

thymoma, these results suggest that the tumours are of clonal origin.

Recombinant nature of new MCF DNA in AKR tumours

We also analysed the DNA of spontaneous tumours for recombinant viral DNA sequences, again using the cell DNA of MCF virus-induced tumours as positive controls for specific DNA fragments. In these studies, we hybridized the restriction endonuclease-digested tumour DNAs to the 3'-viral probe shown below Fig. 3a; this probe recognizes viral sequences to the right of 7.0 kb and includes the *XbaI* site at 7.7 kb which had been found in the DNA of pathogenic MCF viruses but had not been detected in the non-ecotropic sequences of AKR embryo DNA mapped with the BE probe in Fig. 3.

As noted earlier, all MCF viruses contained a *BclI* site at 7.0 kb, as did the endogenous non-ecotropic sequences in embryo DNA. Using the 3' probe, the lack of linkage in AKR embryo DNA of the *BclI* site to an *XbaI* site corresponding to 7.7 kb was confirmed, as embryo DNA digested with *BclI*–*XbaI* did not contain the 0.7-kb *BclI*–*XbaI* internal viral fragment (Fig. 5a, arrow) found in the pathogenic MCF viruses and easily detected in the cell DNA of MCF virus-induced tumours (Fig. 5a, tumour I 01). However, the 0.7-kb *BclI*–*XbaI* fragment was detected in each spontaneous thymoma (Fig. 5a, S 01 and S 02), even in those tumours in which no new viral fragment had been detected with the other probes (tumour S 01). The recombinant 0.7-kb *BclI*–*XbaI* fragment was also present in all metastases, but was absent from uninvolved intestinal tissue. This same recombinant fragment has also been noted in thymomas arising in mice congenic for *Akv-1* or *Akv-2* (ref. 40) and in C58 mice (data not shown). Such results strongly suggest that a recombinational event between the endogenous non-ecotropic *env* (which donated the *BclI* site) and ecotropic *env* (which donated the *XbaI* site) generated a hybrid *env* which corresponds to the structure of the pathogenic MCF viruses. The invariable association of this fragment with tumour DNA suggests that this specific recombination may be an important aspect of the oncological process in AKR tumours. (After this article was submitted, we learned that Quint *et al.*⁴⁶ and W. Herr and W. Gilbert (personal communication) had independently made similar observations in AKR tumours.)

In attempts to define further the region of recombination, we took advantage of the previously noted finding that many of the endogenous sequences which hybridized with the BE probe contained a *PstI* site at 7.5 kb, while this site was not present in AKR ecotropic MuLV. If the recombination took place to the left of this *PstI* site, then *BclI* (at 7.0 kb)–*PstI* (at 8.2 kb), digestion should yield a 1.2-kb fragment; this was the case for all five spontaneous tumours studied (Fig. 5b, arrow). This result indicates either that recombination occurred between MCF sequences located to the right of 7.0 kb and ecotropic sequences located to the left of 7.5 kb, or that the recombination involved an unusual non-ecotropic *env* which lacked the *PstI* site at 7.5 kb.

We also determined that the non-ecotropic sequences extended leftward at least to the *BamHI* site at 6.2 kb, which had been noted in many endogenous non-ecotropic fragments mapped in embryo DNA. Hybridization of *BamHI*–*XbaI*-digested tumour DNA to the 3' probe or the BE probe, which should reveal a novel 1.5-kb fragment if the *BamHI* site at 6.2 kb were now linked to the *XbaI* site at 7.7 kb, showed that this fragment was present in most (S 01 and S 22) but not all (S 16) spontaneous tumours (Fig. 5c, arrow; only 3' probe results are shown). These results suggested that the non-ecotropic sequences extended leftward at least to 6.2 kb in most of these recombinant DNA fragments, as was also true for the pathogenic MCF viruses.

The mechanism by which the viral recombination has occurred is unknown although it probably follows infection of non-target cells by ecotropic virus. Relatively good homology to the left of 6.2 kb and to the right of 7.3 kb has been noted between ecotropic, xenotropic and MCF viruses, and the non-ecotropic sequences of MCF viruses usually include at least this 1-kb DNA segment^{20,47}. It therefore seems reasonable to speculate that the non-ecotropic *env* of MCF virus genomes has been generated by homologous recombination through a double-crossover mechanism.

MCF virus production in tumours

As noted above, tumours induced experimentally by inoculation with MCF virus contained several new MCF viral DNA fragments, whereas the DNA of spontaneous tumours contained few new viral DNA fragments. We therefore sought to determine whether these biochemical differences would be reflected biologically in the frequency of virus-positive cells in MCF virus-induced tumours versus spontaneous tumours. By infectious centre assay (Table 1), no more than 0.01% of the thymoma cells from spontaneous tumours were releasing infectious virus whereas a much higher proportion of the cells from tumours induced by experimental inoculation with an MCF virus released MCF virus. These results therefore correlate with the biochemical data.

The mechanism underlying the low MCF virus production in spontaneous tumours remains to be elucidated. One attractive hypothesis is that the MCF viral sequences in the spontaneous AKR tumour cells is defective for replication. This mechanism has recently been noted in avian leukosis virus-induced bursal tumours^{44,45}. Alternatively, the virus may be unable to be expressed efficiently in tumour cells. However, the release of MCF virus from a significant proportion of the cells in tumours induced experimentally by MCF247 does not support this conjecture.

Significance of recombinant viral sequences

The presence in spontaneous tumours of recombinant *env* sequences with a structure similar to that found in pathogenic MCF viral DNAs, combined with the lack of new ecotropic viral DNAs in the tumours, suggests that MCF viruses with recombinant *env* represent a more immediate inducer of spontaneous AKR lymphomas. The consistent association of the tumours with sequences present in pathogenic MCF viruses (and their absence from non-tumour tissue) suggests that this specific rearrangement of endogenous sequences is an important step in the development of the spontaneous tumours, perhaps by altering the differentiation state or the proliferative capacity of the target cell directly⁴⁸ or by promoter insertion⁴⁹. The non-pathogenic nature of MCF viruses that have apparently acquired the endogenous MCF-like *env* without also retaining the ecotropic virus-derived sequences around the *XbaI* site at 7.7 kb suggests that activation of those endogenous non-ecotropic sequences is not by itself sufficient for tumour induction. Rather, pathogenicity occurs only after certain specific recombinational events between ecotropic virus and the endogenous MCF-like viral sequences. We presume that the non-ecotropic sequences facilitate adsorption and penetration of the virus into the target tissue. The function of the ecotropic sequences around the *XbaI*

site is not known. However, they may also be important for infection of the target tissue, as pathogenic MCF viruses replicate better in the thymus than do the non-pathogenic viruses we have studied¹⁹.

We thank Drs Alan Rein and Michael Potter for a critical

Received 6 August; accepted 27 October 1981.

1. Gross, L. *Oncogenic Viruses* 2nd edn (Pergamon, Oxford, 1970).
2. Gelb, L. D., Milstein, J. B., Martin, M. A. & Aaronson, S. A. *Nature new Biol.* **244**, 76–79 (1973).
3. Chattopadhyay, S. K., Lowy, D. R., Teich, N. M., Levine, A. S. & Rowe, W. P. *Cold Spring Harb. Symp. quant. Biol.* **39**, 1085–1101 (1974).
4. Stephenson, J. R. in *Molecular Biology of RNA Tumour Viruses* (ed. Stephenson, J. R.) 245–297 (Academic, New York, 1980).
5. Kawashima, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **73**, 4680–4684 (1976).
6. Hartley, J. W., Wolford, N. K., Old, L. J. & Rowe, W. P. *Proc. natn. Acad. Sci. U.S.A.* **74**, 789–792 (1977).
7. Fischinger, P. J., Dunlop, N. M. & Blevins, C. S. *J. Virol.* **26**, 532–535 (1978).
8. Troxler, D. H., Yuan, E., Linemeyer, D., Ruscetti, S. & Scolnick, E. M. *J. exp. Med.* **148**, 639–653 (1978).
9. Troxler, D. H., Lowy, D., Howk, R. R., Young, H. & Scolnick, E. M. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4671–4675 (1977).
10. Rommelaere, J., Faller, D. V. & Hopkins, N. *Proc. natn. Acad. Sci. U.S.A.* **75**, 495–499 (1978).
11. Furth, J., Seibold, H. R. & Rathbone, R. R. *Am. J. Cancer* **19**, 521–604 (1933).
12. Rowe, W. P. *Cancer Res.* **33**, 3061–3068 (1973).
13. Rowe, W. P. *J. exp. Med.* **136**, 1272–1285 (1972).
14. Chattopadhyay, S. K., Rowe, W. P., Teich, N. M. & Lowy, D. R. *Proc. natn. Acad. Sci. U.S.A.* **72**, 906–910 (1975).
15. Chattopadhyay, S. K., Lander, M. R., Rands, E. & Lowy, D. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5774–5778 (1980).
16. Steffen, D. L., Bird, S. & Weinberg, R. A. *J. Virol.* **35**, 824–835 (1980).
17. Rowe, W. P. & Pincus, T. *J. exp. Med.* **135**, 429–436 (1972).
18. Lilly, F., Duran-Reynals, M. & Rowe, W. P. *J. exp. Med.* **141**, 882–889 (1975).
19. Cloyd, M. W., Hartley, J. W. & Rowe, W. P. *J. exp. Med.* **151**, 542–552 (1980).
20. Chattopadhyay, S. K., Lander, M. R., Gupta, S., Rands, E. & Lowy, D. R. *Virology* **113**, 465–483 (1981).
21. Rands, E., Lowy, D. R., Lander, M. R. & Chattopadhyay, S. K. *Virology* **108**, 445–452 (1981).
22. Stephenson, J. R., Tronick, S. R. & Aaronson, S. A. *Cell* **6**, 543–548 (1975).
23. Jamjoom, C. A., Naso, R. B. & Arlinghaus, R. B. *Virology* **78**, 11–34 (1977).
24. Karshin, W. L., Arcement, L. J., Naso, R. B. & Arlinghaus, R. B. *J. Virol.* **23**, 787–798 (1977).
25. Donoghue, D. J., Rothenberg, E., Hopkins, N., Baltimore, D. & Sharp, P. A. *Cell* **14**, 959–970 (1978).

reading of the manuscript, Dr Wallace Rowe for tumour tissues from *Akv* congenic mice, and Mrs Lorraine Shaughnessy and Mr Harry Schaefer for preparation of the manuscript. This research was supported in part by a contract from the National Cancer Institute to Meloy Laboratories.

26. Rothenberg, E., Donoghue, D. J. & Baltimore, D. *Cell* **13**, 435–451 (1978).
27. Sutcliffe, J. G., Shinnick, T. M., Verma, I. M. & Lerner, R. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3302–3306 (1980).
28. Lung, M. L., Hering, C., Hartley, J. W., Rowe, W. P. & Hopkins, N. *Cold Spring Harb. Symp. quant. Biol.* **44**, 1269–1279 (1979).
29. Linemeyer, D. L., Ruscetti, S. K., Menke, J. G. & Scolnick, E. M. *J. Virol.* **35**, 710–721 (1980).
30. Ruscetti, S., Linemeyer, D., Feild, J., Troxler, D. & Scolnick, E. J. *exp. Med.* **148**, 654–663 (1978).
31. Astrin, S. M. *et al. Cold Spring Harb. Symp. quant. Biol.* **44**, 1105–1110 (1979).
32. Elder, J. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **74**, 4676–4680 (1977).
33. Green, N. *et al. J. exp. Med.* **152**, 249–264 (1980).
34. Stockert, E., DeLeo, A. B., O'Donnell, P. V., Obata, Y. & Old, L. J. *J. exp. Med.* **149**, 200–215 (1979).
35. Cloyd, M. W., Hartley, J. W. & Rowe, W. P. *J. exp. Med.* **149**, 702–712 (1979).
36. O'Donnell, P. V., Stockert, E., Obata, Y., DeLeo, A. B. & Old, L. J. *Cold Spring Harb. Symp. quant. Biol.* **44**, 1255–1264 (1979).
37. Niman, H. L. & Elder, J. H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4524–4528 (1980).
38. Pedersen, F. S., Buchhagen, D. L., Chen, C. Y., Hays, E. F. & Haseltine, W. A. *J. Virol.* **35**, 211–218 (1980).
39. Canaan, E. & Aaronson, S. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1677–1681 (1979).
40. Steffen, D. L., Bird, S., Rowe, W. P. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4554–4558 (1979).
41. Yoshimura, F. K. & Breda, M. J. *J. Virol.* **39**, 808–815 (1981).
42. Gisselbrecht, S., Blaineau, C., Hurot, M. A., Pozo, F. & Levy, J. P. *Cancer Res.* **38**, 939–941 (1978).
43. Blank, K. J. & Pincus, T. *J. exp. Med.* **152**, 458–462 (1980).
44. Payne, G. S., Neel, B. G., & Astrin, S. M. *Nature* **290**, 475–480 (1981).
45. Neel, B. G., Hayward, W. S., Robinson, H. L., Fang, J. & Astrin, S. M. *Cell* **23**, 323–334 (1981).
46. Quint, W., Quax, W., van der Putten & Berns, A. *J. Virol.* **39**, 1–10 (1981).
47. Chien, Y., Verma, I. M., Shih, T. Y., Scolnick, E. M. & Davidson, N. J. *J. Virol.* **28**, 352–360 (1978).
48. McGrath, M. S. & Weissman, I. L. *Cell* **17**, 65–75 (1979).
49. Hayward, W. S., Neel, B. G., & Astrin, S. M. *Nature* **290**, 475–480 (1981).
50. Lowy, D. R., Rands, E., Chattopadhyay, S. K., Garon, C. F. & Hager, G. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 614–618 (1980).
51. Hirt, B. *J. molec. Biol.* **26**, 365–369 (1967).
52. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).

The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences

Graeme I. Bell, Mark J. Selby & William J. Rutter

Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, USA

The sequence of the highly polymorphic region near the human insulin gene, which begins 363 base pairs before the start of transcription and extends upstream, indicated that it is generated by variation in the number and arrangement of members of a family of tandemly repeating nucleotides whose structure is related to ACAGGGGTGTGGGG. Another 15-base pair sequence, repeated twice near the human gene, is located ~325 base pairs on either side of the polymorphic region. One member is in the transcriptional control region.

THE human insulin gene region of two human chromosome 11 homologues has been isolated and characterized^{1–5}. The organization and structure of the 1,430-base pair (bp) segment encoding the insulin mRNA precursor and ~40 kilobases (kb) of the flanking sequence on different chromosomes are similar, but restriction mapping studies of 150 individuals (ref. 6 and G.I.B. and J. Karam, unpublished results) indicated that a highly polymorphic region is located ~500 bp before the point where transcription of insulin mRNA begins⁶. Sixty-three per cent of non-diabetic caucasian individuals examined are heterozygous at this site, and there are two major and one minor class of alleles⁶. However, as there is heterogeneity within each class, there are many different alleles at this locus. Preliminary data also indicate that there may be racial differences in the frequency of alleles, that this polymorphism is inherited in a mendelian fashion in a four-generation pedigree and that the organization of the polymorphic region in DNA prepared from

an insulin-secreting adenoma and white blood cells of the same individual is identical (G.I.B., unpublished). However, the molecular events which produce either this highly polymorphic region or one with similar properties that has been described at an unidentified locus⁷, have not been determined.

Restriction mapping studies of the insulin gene region suggested that the restriction fragment length polymorphism is generated by the insertion–deletion of DNA sequences, so that fragments of different size are generated when DNA from a heterozygous individual is digested with selected restriction endonucleases. The insertions in this region are primarily of two size classes: 0–600 bp and 1,600–2,200 bp; restriction fragments possessing insertions of 600–1,600 bp are rarely observed. We have sequenced the corresponding polymorphic regions of two chromosomes to determine the basis for this length polymorphism. The sequence comparison indicated that the polymorphism is generated by variation in the redundancy of

GGGGT and GGGG. Both nucleotide substitutions and single nucleotide insertions are observed: G → CC, G → C, GT → CCC. And in addition the boundaries of this region, 5'CTGGGG,ACAG3', are found in the 14-bp sequence.

Although the arrangement of the various oligonucleotide sequences within the polymorphic regions is different (Fig. 2b), the composition and order of the three units at the 5' boundary and the four units of the 3' boundary of this region are identical. Moreover, the variant sequence ACAGGGTCCTGGGG (in Table 2) is found only once in both arrays in this conserved 5' boundary region. These data suggest that the molecular processes which generate the divergence within this family operate on the central regions of the array.

Length variation of polymorphic forms

We have previously shown that the polymorphism exists in two major forms⁶. Approximately 75% of all the insulin gene regions analysed have a short polymorphic region (0–600 bp) and ~25% possess a longer polymorphic region (1,600–2,200 bp). Polymorphic regions of intermediate size are rarely observed (6/300 chromosomes examined). The two allelic DNA segments which we have sequenced are of the more frequent shorter form. As the length heterogeneity in this class is essentially due to variation in the number of oligonucleotide repeats, our previous data⁶ can be reinterpreted. The shorter polymorphic region is composed of alleles with 26–63 repeating oligonucleotide units, with an average of 38 repeats.

We have examined the nature of the longer polymorphic form by hybridizing an *in vitro* labelled 879-bp *Pvu*II fragment of the flanking region of allele 2 (AH2, the allelic DNA segment corresponding to the region at nucleotides 1,206–1,927, Fig. 1) to a *Sac*I digest of DNA of nine individuals, four of whom possessed both a short and a long polymorphic region (individuals 019, 025, 035 and 038, Fig. 2). This 879-bp *Pvu*II fragment included 45 repeating units of 641 bp, as well as 134 bp on the 5' end and 105 bp on the 3' end. We reasoned that if the length variation in the larger polymorphic region was due to this same repeating oligonucleotide, we should observe more intense hybridization to DNA fragments containing the longer polymorphic region.

There was more intense hybridization to the larger fragment in four individuals: 019, 025, 035 and 038 (Fig. 2). In contrast, a DNA fragment specific for the insulin gene, which included a 59-bp 5'-flanking sequence, 1,430-bp structural region and 118-bp 3'-flanking sequence, hybridized with equal intensity to both DNA fragments (data not shown and Fig. 2C in ref. 6). Thus, this differential autoradiographic signal is not due to

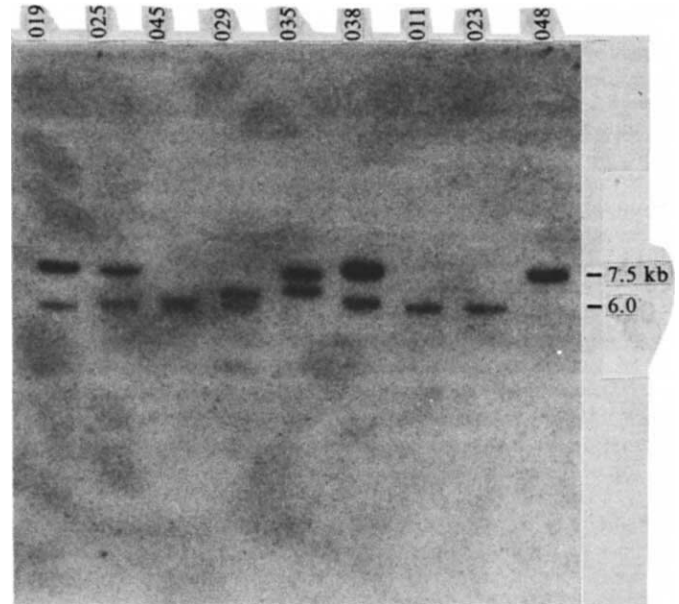


Fig. 2 Hybridization of a DNA fragment containing the region of tandemly repeating oligonucleotides to *Sac*I digests of human DNA. DNA was prepared from white blood cells, digested with *Sac*I, the fragments were separated by electrophoresis in a 1% agarose gel, transferred to nitrocellulose filters and hybridized as described previously⁶. Individuals are indicated by number above the autoradiogram and sizes of hybridizing fragments on the right.

differential transfer or selective loss of DNA from the filter but is due at least in part to variation in the redundancy of this family of oligonucleotides. However, without cloning and sequencing the longer polymorphic allele or the rare intermediate form, we cannot exclude the possibility that another sequence in the ~100 bp on either side of the repeating oligonucleotides has not been selectively amplified. There are no other tandemly repeated sequences flanking the polymorphic region, so if additional sequence amplification occurred, it is of a non-repeated sequence. However, as we have defined the boundaries of the region of length variation for all forms, *Bgl*II and *Pvu*II sites at nucleotides 1,025 and 1,927 (Fig. 1), respectively^{2,6}, and more recently between *Pvu*II sites at 1,206 and 1,927 (unpublished), we believe that the variation in the redundancy of this family of G+C-rich oligonucleotides accounts for all length variation. Thus, the longer polymorphic regions contain 141–209 repeats, with an average of 163. The rare intermediate forms possess ~80 repeats and the short form possesses ~40 repeats. The significance of the arithmetic progression in the number of repeats is unknown.

Presence of the polymorphic region in the genome

We have been unable to detect additional regions of hybridization of the 879-bp *Pvu*II fragment to digests of genomic DNA (Fig. 2) or the ~45 kb of DNA sequence around and including the insulin gene (data now shown). Thus, the sequences within the 879-bp *Pvu*II fragment are restricted to the region flanking the 5' end of the insulin gene or at least they are not tandemly repeated elsewhere in the genome. This confirms our previous results². The redundant oligonucleotide region is therefore a single copy sequence.

Comparison with the rat insulin I gene

We have examined the 5'-flanking region of the rat insulin I gene for a similar polymorphic region and have determined the sequence of this region of the gene from a *Pvu*II site 414 bp from the 5' end to the beginning of the mRNA coding region (Fig. 3). There were no tandemly repeating sequences. If such sequences exist they must be more than 414 bp from the 5' end of the coding region.

Table 1 Non-tandemly repeated sequences of ≥10 bp in the 5'-flanking region of the human insulin gene

Sequence	Positions*
GGCCAGGGTG	60, 684
AGCAGCCCCAG	91, 1,208 rc
GCCTCCCTCC	150, 445
CTCAGCTCCC	194, 1,189 rc
CAGCTGGGGG	323, 1,921 rc
AGGGAGGGGTCC	578, 1,260 rc
AGGGGTCCTG	582, 1,342, 1,567, 1,638, 1,724, 1,767, 1,782, 1,797
CAGGGTGGGA	687, 1,227
CCCTGGTGCT	788, 2,035 rc
CTGGGGAGCTG	884, 1,186
AGGGGAGATGGGCTC	976, 1,135
TGGGGGCTGAG	1,003, 2,067 rc
CCCACCCCA	1,069, 1,282
CCACCCCCAGA	1,070, 1,090
TGTCCTCCAGG	1,147, 1,348 rc, 1,376 rc, 1,573 rc, 1,730 rc, 1,773 rc, 1,788 rc, 1,803 rc
GCTGCTGTCC	1,212, 1,822 rc
GCTCTCTG	1,859, 1,902

* Sequences denoted 'rc' form an inverted repeat.

A comparison of the extended 5'-flanking sequences of the human and rat I insulin genes (Fig. 3) indicated that the homology is more extensive than suggested by the heteroduplex analysis². In fact, interrupted regions of nucleotide sequence homology extend for ~365 bp from the structural portion of the gene, the homology apparently ending at or near the boundary of the polymorphic region of the human gene without being re-established on its other side. This nucleotide sequence conservation suggests that DNA sequences besides the TATAAA (Hogness box) sequence at -25 to -30 and the remote control sequence at -75 to -90 are involved in the regulation of insulin gene expression.

The regions of homology might represent points of interaction between the DNA and other regulatory molecules. Wu and Gilbert¹⁰ have also suggested that additional sequences in the 5'-flanking region of the insulin gene may be involved in the regulation of insulin gene expression. They observed that a region of the rat insulin II gene, 250–300 bp upstream from the start of transcription, exhibits tissue-specific DNase sensitivity. Lomedico *et al.*¹¹ also observed that the homology between the rat insulin I and II genes included ~500 bp in the 5'-flanking region.

Discussion

The human insulin gene and its flanking regions exhibit conserved and highly variable sequences. Comparison of 2,357-bp segments from two chromosomes showed only four substitutions (4/2,357, 0.17%; this paper and ref. 5), all within the 1,430-bp portion encoding the mRNA precursor. The 5'-flanking region contains a section of extensive length and sequence heterogeneity which is generated by variation in the number and arrangement of members of a family of G + C-rich oligonucleotides whose consensus sequence is ACAGGGGTGTGGGG. This polymorphic region, although internally repetitious, is a single-copy structure. A comparison of the size of this region in individuals indicates that it can vary from 26 repeat units (364 bp) to 209 units (2,926 bp). The latter generates a DNA segment twice the size of the structural region (1,430 bp) of the human insulin gene. Moreover, the sequence polymorphism suggests that those alleles which cannot be distinguished by size may be distinguished by sequence.

Table 2 Organization of the oligonucleotide repeats in two allelic polymorphic regions

Sequence	Occurrences	
	Allele 1 (Ahi1)	Allele 2 (Ahi2)
a, ACAGGGGTGTGGGG	13 (38.2%)	24 (53.3%)
b, ACAGGGGTCTGGGG	8 (23.5%)	7 (15.6%)
c, ACAGGGGTCTGGGG	5 (14.7%)	6 (13.3%)
d, ACAGGGGTCCGGGG	2 (5.9%)	1 (2.2%)
e, ATAGGGGTGTGGGG	1 (2.9%)	0
f, ACAGGGGTCCCGGG	1 (2.9%)	5 (11.1%)
g, ACAGGGGTCTGAGG	1 (2.9%)	0
h, ACAGGGGTCTGGGC	1 (2.9%)	0
i, ACAGGGGTCTGGGG	1 (2.9%)	1 (2.2%)
j, ACAGGGGTGTGAGG	0	1 (2.2%)

Arrangement of oligonucleotides

Allele 1 5'cdi aba aba dab aab aca aaa cea faa egh ccc b'3' (ref.45)

Allele 2 5'cdi jfa faa aba baa aaf aba afa baa aac aaa cac baa afc ccb 3' (ref.34)

Upper panel, sequences of different oligonucleotides and the number of times (and percentage of all repeats) they occur in the two allelic DNA segments. Lower panel, arrangement of oligonucleotides in the polymorphic region. Each oligonucleotide sequence was assigned a letter as indicated in the upper panel.

The insulin polymorphic region can be distinguished from other regions of tandemly repeating oligonucleotides. Its unique chromosomal location and single copy nature distinguish it from the hexanucleotide sequences repeated at the ends of the fragmented, macronuclear DNA molecules of some protozoa^{12,13}, while its smaller size, greater internal redundancy and sequence simplicity distinguish it from the repeated sequences found at the *ori* region of SV40 that have been implicated in the regulation of replication and perhaps transcription^{14,15}. The insulin polymorphic region is similar to a region in the *Xenopus* 5S rRNA gene nontranscribed spacer which is also polymorphic and possess 2–12 copies of a 15-bp oligonucleotide¹⁶ as well as to simple tandem repetitive sequences in the recombinational switch regions of immunoglobulin heavy chain genes¹⁷. The highly repetitive satellite DNAs present in eukaryotes are often composed of shorter repeated units (oligonucleotides)^{18,19}. However, the insulin polymorphic region is only found at one site in the human genome, in contrast to those in the rRNA genes and satellite DNAs, which are present at many locations.

The sequence simplicity of the polymorphic region suggests that it could be involved in some recombinational process. However, the persistence of the bimodal distribution of insertion sizes⁶ suggests that random crossing over within this array does not occur frequently because such exchanges would be expected to produce a unimodal distribution²⁰. The insulin polymorphic region is more complex than the repeating T·G dinucleotide region in intervening sequence 2 of the human fetal γ -globin genes which has been implicated in gene conversion²¹. A similar region of repeating C·T dinucleotides has been observed in the sea urchin histone H2A and H1 spacer and it has been proposed that this region is involved in the maintenance of homogeneity of the histone gene repeating unit²². The sequence data on the regions flanking the insulin polymorphic region provide no evidence for gene conversion. There are 0/525 nucleotide differences on the 5' side of the polymorphic region and 4/1,832 on the 3' side (this paper and ref. 5).

The function of this polymorphic region is unknown. It could be intergenic or spacer DNA. Its location, within 363 bp or 2 nucleosomes from the coding region of the gene, adjacent to the putative transcription control region, suggests that the polymorphic region could have a role in the regulation of human insulin gene expression. However, this region does not seem to

```

1781 CAGGGGTCTGGGGACAGGGGTCTGGGGACAGGGGTCTGGGGACAGCAGC CAAAGAG
    *** * * * *
1 CAGCTGAGCTAAGAATCCAGCTATCAATGAAACTATGAAACAGTTCACGGGACAAAGAT
1840 CCCCAGCTGCAGCCTCCAGCTCTCTGCTCTAAT GTGGAAGTGGCCAGGTCTGACGAC
    * * * * *
61 ACCAGGTCCCAACAACTGCACTTTCTGGGAAATGAGGTGGAAATG CTCACCAAGG
1897 GCTTGTCTCTCTGGAGACATTTGCCCCAGCTGTGAGCAGGGACAGG TCTGGCC
    * * * * *
120 AAAAAAGAGGGCCTTACCTCTCTGGGACAAATGATTGTGTGTAAGTCTTCATCAGGCC
1952 ACCGGGCCCCCTGGTTAAGACTCTAATGACCCGCTGGTCTGAGGAAGAGGTGCTGACGAC
    * * * * *
180 ATCTGGCCCTTG TTAATACTTAATTACCTA GGTC TAAGT AGAGTGTGTGACGTC
2012 CAAGGAGATCTTCCACAGACCCAGCAGGGAATGGTCCGGAATTCAGCCTCAGC
    *** * * * *
236 CAATGAGCGCTTCTGACAGCTAGCAGTGGCAAGTG TTTGGAATTCAGCTTCAGC
2072 CCCCC GCATCTGCGGACCCGCCACCCC GCCCT AATGGGCCAGGCGGAGGG
    *** * * * *
295 CCTCTCGCATCTGCTACCTACCTCTCTAGAGCCCTTAATGGGCCAAAGCGAAAGT
2126 TTGACAGGTAGGGAGATGGGCTCTGAGACTATAAGCCAGCGGGGCCAGCAGCCCTC
    *** * * * *
355 CCAGGGGCGAGAGAGGAGGTGCTTTG GACTATAAGCTAGTGGAGACCCAGTAACCTCC
2186 AGCCCTCCAGGACAGGCTGCA TCAGAGAGGCCATCAAGCAG
    * * * * *
414 AACCTAAGTGACAGCTCAATCATAGACCATCAGCAAGCAG
  
```

Fig. 3 Sequence of the 5' flanking region of the rat insulin gene and homology with the corresponding region of the human insulin gene. The rat I and human insulin gene sequences are indicated. The sequence of coding region 1 is also included. The putative 5' end of the mRNA is A-414 in the rat I sequence and A-2186 in the human sequence. The isolation and characterization of the rat insulin I gene has been described²⁸. The sequence was determined by the procedure of Maxam and Gilbert¹⁰ as described in Fig. 1 legend. The human sequence is from Figure 1. Asterisks indicate homologous nucleotides. Gaps have been used to maximize homology.

affect the transcription of the gene in cell-free extracts prepared from HeLa cells (G.I.B. and O. Laub, unpublished). If similar regions of tandemly repeating sequences have a general role in the regulation of insulin expression, they might be found flanking insulin genes of other species. Our sequence analysis of the rat insulin I gene indicated that it does not possess an homologous region in the same position, and that if present it is more than 50 bp upstream from the corresponding position of the human gene. However, this does not preclude a role for this region, as there are two non-allelic insulin genes in rats and only one in humans and most other organisms²³. Thus, some elements of regulation of gene expression could be different.

The role of the additional, non-tandemly repeated sequences in the flanking region of the human gene is also unknown. The 15-bp sequence (Table 1) which is directly repeated ~325 bp on either side of the polymorphic region is particularly interesting as one member is located in the control region of the gene. It could be involved in the function of the polymorphic region or define some aspect of expression of the insulin gene. Though far

and variably distant in the DNA sequence, these two sites could be contiguous in chromatin. Physical joining of these sites in chromatin could affect insulin gene transcription by regulating access of the RNA polymerase to the control regions.

The analysis of length variation in the 5'-flanking region of the human insulin gene indicated that 63% of the individuals examined were heterozygous at this site⁶. As this region also possesses DNA sequence polymorphism, the extent of heterozygosity was underestimated and it may be possible to distinguish the two parental alleles in most individuals by analysing this region for both length and sequence differences. This polymorphic region is a useful molecular marker for the insulin gene on the short arm of chromosome 11 (refs 4, 24) and may serve as marker for loci up to 10–20 × 10⁶ bp distant²⁵.

We thank Dr D. N. Standing for useful discussions and Ms L. Spector and S. Bock for assistance in the preparation of this manuscript. This research was supported by an Elliott P. Joslin Research and Development Award of the American Diabetes Association, Inc., to G.I.B., and NIH grants to W.J.R.

Received 20 August; accepted 3 November 1981.

1. Bell, G. I. *et al. Nature* **284**, 26–32 (1980).
2. Bell, G. I., Pictet, R. L. & Rutter, W. J. *Nucleic Acids Res.* **8**, 4091–4109 (1980).
3. Owerbach, D., Bell, G. I., Rutter, W. J. & Shows, T. B. *Nature* **286**, 82–84 (1980).
4. Owerbach, D., Bell, G. I., Rutter, W. J., Brown, J. A. & Shows, T. B. *Diabetes* **30**, 267–270 (1981).
5. Ullrich, A., Dull, J. J., Gray, A., Brosius, J. & Sures, I. *Science* **209**, 612–615 (1980).
6. Bell, G. I., Karam, J. H. & Rutter, W. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5759–5763 (1981).
7. Wyman, A. R. & White, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6754–6758 (1980).
8. Corden, J. *et al. Science* **209**, 1406–1414 (1980).
9. Elstratidis, A. *et al. Cell* **21**, 653–668 (1980).
10. Wu, C. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1577–1580 (1981).
11. Lomedico, P. *et al. Cell* **18**, 545–558 (1979).
12. Klobutcher, L. A., Swanton, M. T., Donini, P. & Prescott, D. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3015–3019 (1981).
13. Katzen, A. L., Cann, G. M. & Blackburn, E. H. *Cell* **24**, 313–320 (1981).
14. Benoist, C. & Chambon, P. *Nature* **290**, 304–310 (1981).

15. Gruss, P., Dhar, R. & Khoury, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 943–947 (1981).
16. Fedoroff, N. *Cell* **16**, 697–710 (1979).
17. Nikaide, T., Nakai, S. & Honjo, T. *Nature* **292**, 845–848 (1981).
18. Brutlag, D. L. *A. Rev. Genet.* **14**, 121–144 (1980).
19. Streeck, R. E. *Science* **213**, 443–445 (1981).
20. Smith, G. P. *Science* **191**, 528–535 (1976).
21. Slightom, J. L., Blechl, A. E. & Smithies, O. *Cell* **21**, 627–638 (1980).
22. Sures, I., Lowry, J. & Kedes, L. H. *Cell* **15**, 1033–1044 (1978).
23. Humbel, R. E., Bosshard, H. R. & Zahn, H. in *Handbook of Physiology* Sect. 7, Vol. 1 (eds Steiner, D. F. & Freinkel, N. 111–132 (American Physiological Society, Washington DC), 1972).
24. Harper, M. E., Ullrich, A. & Saunders, G. F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4458–4460 (1981).
25. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. *Am. J. hum. Genet.* **32**, 314–331 (1980).
26. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
27. Cooke, N. E., Coit, D., Shine, J., Baxter, J. D. & Martial, J. A. *J. biol. Chem.* **256**, 4007–4016 (1981).
28. Cordell, B. *et al. Cell* **18**, 533–543 (1979).

Lysis gene expression of RNA phage MS2 depends on a frameshift during translation of the overlapping coat protein gene

R. A. Kastelein, E. Remaut*, W. Fiers* & J. van Duin†

Department of Biochemistry, University of Leiden, PO Box 9505, 2300 RA Leiden, The Netherlands

* Laboratory for Molecular Biology, University of Ghent, Ghent, Belgium

Translation of the lysis cistron of MS2 RNA requires a frameshift of ribosomes reading the upstream coat protein mRNA. The out-of-phase fraction of ribosomes terminates at either one of two nonsense codons just preceding the lysis cistron. Termination is followed by initiation at the start codon of the lysis protein message.

THE group I single-stranded RNA bacteriophages (including MS2, f2, R17 and M12) contain the information for at least four proteins^{1–5}, the order of which is illustrated in Fig. 1. The figure shows that the newly characterized L gene, whose product is involved in lysis of the host, overlaps the coat and synthetase gene in the +1 reading frame. The complete primary sequence of MS2 RNA has been published previously⁶.

The timing and efficiency of phage gene expression in the infected cell has been extensively reviewed for the three longer known genes⁷. As for the lysis gene, why has it evolved in its present position, where there is overlap with two other genes? In particular, in what way does this position relate to the level and

time of expression? The L protein is made in small amounts and its appearance in the infected cell coincides with that of the coat protein⁵. A relationship between coat and L protein synthesis is suggested by the observation that phage coat ambers do not lyse their (non-permissive) host⁸.

The availability of cloned MS2 DNA⁹ allowed us to examine how the expression of the L gene is regulated. We used the recombinant DNA approach, which has several advantages over studies using normal phage infection: it allows the site-selected introduction of almost any kind of mutation and the expression of a single gene can be examined.

Our results indicate that the lysis cistron does not possess an independent entry site for ribosomes. Instead its translation depends on the readout of the coat gene. Ribosomes that experience a frameshift when reading the coat gene preceding the lysis cistron, terminate protein synthesis a few codons before

† Present address: Thimann Laboratories, University of California, Division of Natural Sciences, Santa Cruz, California 95064, USA.

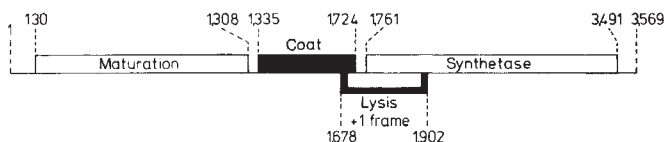


Fig. 1 Genetic map of the RNA phage MS2. Nucleotide numbers are taken from ref. 6. The reading frame of the lysis gene is +1 with respect to the coat and synthetase frame.

this gene due to the presence of two out-of-phase stop codons. Termination of protein synthesis at these positions elicits synthesis of the lysis protein.

Expression of the lysis gene in wild-type MS2 DNA subgenomic regions

In this study, selected regions of the MS2 genome were inserted into plasmid pPLa831, just downstream of the λ leftward promoter pL. The plasmid has been described in detail elsewhere¹⁰. The promoter pL is closed at 28 °C by a thermosensitive repressor encoded by the host chromosome. At 42 °C the repressor is unstable and the promoter is turned on.

The stringency of temperature control over the promoter in pPLa831-derived plasmids is demonstrated in Fig. 2. (Here, *Escherichia coli* strain M5219 (ref. 10) was transformed with pMS2 (Table 1) and labelled with radioactive amino acids.) At 28 °C, no MS2 coat protein is detected (Fig. 2, lanes 1 and 2), but at 42 °C the coat protein of the phage becomes the major product synthesized (lanes 3–7). Similarly, the presence of the almost complete MS2 DNA genome (nucleotides 103–3,569) in pPLa831 did not interfere with the growth of the transformed cells as long as transcription from pL was blocked by keeping the culture at 28 °C. However, if this particular clone (pMS23) was shifted to 42 °C, complete lysis was observed (Fig. 3a and Table 1). Apparently the L gene is expressed in the experimental conditions used. Evidence that cell lysis can indeed be ascribed to the presence of the L gene is provided by a clone containing pMS20. This plasmid carries a 58 nucleotide-long deletion in the L gene (Table 1). As shown in Fig. 3a, the deletion abolishes cell lysis, in agreement with the proposed genetic map.

To determine a possible contribution of the other gene products to cell lysis we deleted large parts of the maturation protein gene and the synthetase gene to obtain plasmid pMS16. Cells harbouring this plasmid still lyse on induction of the pL promoter (Fig. 3a). Thus neither the maturation (A) protein nor the synthetase appears necessary for cell lysis. The ability of plasmids to lyse the host cell is maintained up to a point where the whole A protein is removed (pMS17; Table 1).

L gene expression requires overlap of coat and L genes

When the MS2 DNA information in pMS17 was trimmed away further to give pMS1 (Table 1), there was no cell lysis at 42 °C (Fig. 3b). There are several possible explanations for this and we have considered the two that are more likely. (1) L protein synthesis is under translational polar control of the coat gene. In this case, ribosomes reading the coat message must open up the initiation region of the L cistron to allow independent initiation by another ribosome. The start region of the L cistron is thus thought to be inaccessible to ribosomes in the unperturbed mRNA. In ribosome binding experiments, this region of MS2 RNA has been found to be unprotected against RNase digestion^{4,11}. (2) Absence of cell lysis in pMS1 clones could indicate that the coat protein itself is in some way required for cell lysis to occur—involvement of the coat protein has been suggested previously⁸.

To test alternative (2), we modified pMS1 by inserting the MS2 DNA region specifying the coat protein behind the L gene, thus uncoupling the gene overlap. Clones carrying this new plasmid (pMS1.9) did not lyse on induction, although the presence of the MS2 coat protein was evident from protein gel

analysis of the type shown in Fig. 2. This suggests that L protein synthesis depends on the native gene arrangement and supports a polar control mechanism. Further evidence for the involvement of translational polar control was provided by pMS25. Here we deleted the initiation region of the coat gene and the ability of this plasmid to induce cell lysis was lost (Table 1; and Fig. 3b). As before, lysis is not restored when the information for the coat protein is supplied behind the L gene (pMS25.9). Note that in pMS25 the 5' context of the L gene is kept intact for over 300 nucleotides. This virtually excludes the possibility that the lack of mRNA translation is due to the deletion of translational start signals (in the conventional sense).

Efficiency of coat protein synthesis correlated with onset of cell lysis

During this study we obtained independent evidence that the translation of coat and L protein mRNA sequences is coupled; clones producing large amounts of coat protein also lysed quickly. In contrast, clones synthesizing small amounts of coat protein showed delayed lysis. Variations in coat gene expression can be obtained by varying the length of the MS2 DNA region proximal to the coat gene. For example, after induction of pMS2, coat protein comprised 30% of the newly synthesized proteins after 30 min, as opposed to <5% for pMS17. Cell lysis (as measured by the decrease in turbidity) began after 30 min for pMS2 and only after 75 min for pMS17. The data for several clones are summarized in Table 2. A positive correlation between the rate of MS2 coat protein synthesis and the onset of cell lysis is apparent. Such a relationship is consistent with a polar translational control mechanism.

Fine regulation of polar control

During phage infection, polar control of synthetase translation is observed in coat protein mutants that have a nonsense codon in the position of the sixth amino acid. The control is relieved when the stop codon is found further downstream at amino acid residue 50, 54 or 70 (ref. 12). Based on these data and the MS2 RNA secondary structure model, a molecular basis for this polarity has been proposed¹³, the main feature of which is the base-pairing between the synthetase initiation region and a coat cistron coding sequence (see Fig. 4). Ribosomes passing this region are thought to expose the synthetase ribosomal binding site to allow independent initiation of protein synthesis. A similar situation may exist at the start of the L gene. Assuming the validity of the 'coat-lysis polarity control' one would expect that the L gene becomes readable once ribosomes pass over

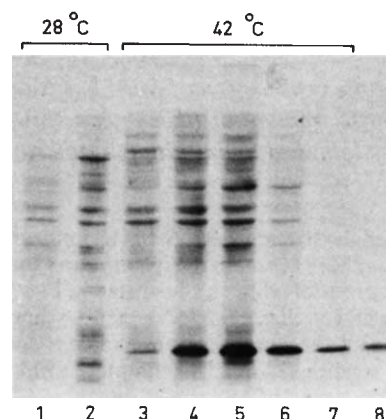


Fig. 2 Expression of MS2 DNA in cells transformed with pMS2. Cultures were grown to 0.2 A_{650} , then labelled with 14 C-amino acids for the time intervals indicated below. Proteins were analysed on polyacrylamide gels containing 0.1% SDS. Lanes 1, 2: pattern after labelling at 28 °C for 5–25 and 95–115 min, respectively. Lanes 3–7: labelling at 42 °C for 5–10, 5–25, 35–55, 65–85 and 95–115 min, respectively. Lane 8: *in vitro*-labelled MS2 coat protein.

Table 1 Host cell lysis by pPLa831-derived plasmids after induction at 42°C

Plasmid no.	MS2 DNA inserted in pPLa831	Cell lysis after induction at 42°C	Schematic representation of MS2 genes present
pMS1	1,628-2057	-	
pMS1.9	1,628-2057, 1,221-1,736	-	
pMS2	103-2057	+	
pMS9	1,221-1,736	-	
pMS12	1,221-2057	+	
pMS16	869-2057	+	
pMS17	1,305-2057	+	
pMS18	869-3569	+	
pMS20	869-2057 (∇ 1,764-1,822)	-	
pMS23	103-3569	+	
pMS25	869-2057 (∇ 1,013-1,365)	-	
pMS25.9	869-2057 (∇ 1,013-1,365), 1,221-1,736	-	
MS2 DNA inserted in pPLa932			
pMS32.00	1,221-2057 1,628 1,629	+	
pMS32.04	1,221-2057 (G...AAT...TAA...TTC.)	-	
pMS32.259	1,221-2057 (∇ 1,369-1,628) GTC...GAA...ATT...C	+	
	1,369 1,629		

All nucleotide numbers refer to the primary structure of MS2 RNA (Fig. 1; ref. 6). MS2 DNA was inserted in pPLa831 according to standard procedures using either the *EcoRI* site, the *BamHI* site or both. When the MS2 DNA region of interest did not contain a suitable site, *EcoRI* or *BamHI* restriction site linker sequences were attached. MS2 DNA sites used were: *EcoRI* (103) (nucleotide number), *PvuII* (869), *SalI* (1,013), *HaeIII* (1,221), *XbaI* (1,305), *SalI* (1,366), *EcoRI* (1,628), *HaeIII* (1,736), *BamHI* (2,057). pPLa932 is identical to pPLa831 except that the *EcoRI* site was removed by cutting with *EcoRI*, filling in the single-stranded regions with DNA polymerase I (large fragment) and re-sealing with T4 DNA ligase. Thus, pMS32.00 is equivalent to pMS12 except for the *EcoRI* site in the vector. pMS32.04 was constructed by opening up the *EcoRI* site of pMS32.00 at position 1,628. The resulting sticky ends were filled with DNA polymerase I (large fragment) and the plasmid re-sealed with T4 DNA ligase. The new sequence around 1,628 is shown. The coat reading frame is indicated by dots. The construction of pMS32.259 is described in the text. The distance between the *SalI* and *EcoRI* positions is 263 nucleotides. As 4 nucleotides are recovered by the action of the polymerase, the final deletion is 259 nucleotides. The reading frame is indicated by broken lines.

nucleotides 1,563-1,570. To test this, we introduced a UAA stop signal in the coat gene beyond this region, that is, at position 1,632. Surprisingly, cells harbouring this new plasmid, pMS32.04, did not lyse the host cell (Fig. 3c). The expression of the inserted MS2 DNA was apparent from the synthesis of a protein having the size of the expected coat protein ochre fragment (Fig. 6). The absence of lysis in clone pMS32.04 indicated that for L cistron translation to occur, ribosomes have to move beyond position 1,632 of the MS2 RNA sequence. To achieve this we again modified pMS32.00. The unique *SalI* and *EcoRI* restriction sites, at positions 1,366 and 1,628 respectively, were used to delete this region of 263 nucleotides. The resulting staggered ends were filled in with DNA polymerase I (large fragment) and the plasmid re-sealed (pMS32.259). A consequence of the deletion was a shift to the +1 reading frame at the junction site. The +1 reading frame allows ribosomes that start at the coat cistron to proceed up to position 1,672, where they encounter an ochre stop codon, three nucleotides before the L cistron start (Fig. 4). Clones containing this plasmid (pMS32.259) lysed quickly on induction. Their growth curve was almost identical to that of pMS32.00 (not shown). From this we make two important conclusions. First, intact coat protein is not required *per se* for the synthesis or activity of the L protein. Second, ribosomes shifting their reading frame to +1 during coat cistron translation, so as to arrive at the ochre stop codon at nucleotide 1,672, trigger synthesis of the L protein.

Reading frameshifts in coat cistron promote synthesis of lysis protein

Results from clone pMS32.259 suggest that synthesis of the lysis protein from wild-type RNA may depend on a ribosomal frameshift to the +1 mode at the appropriate region. To explore this further, we constructed a collection of coat gene deletion

mutants that potentially contain all three reading frames. Experimentally, pMS32.00 was opened at its unique *EcoRI* site and the linearized DNA treated with exonuclease *Bal31* for 15, 30, 60 and 90 s, respectively. The reaction mixtures were pooled, the DNA ligated and used for transformation. Of 250 recombinants obtained, 60 were selected for further restriction analysis and lytic properties. The growth curves obtained were of three general types: quick lysis, delayed lysis and no lysis.

Nine clones, with the general nomenclature pMS32.N, where N is the number of nucleotides deleted, were selected for sequence analysis of the modified region (Table 3). Clones pMS32.16 and pMS32.28 change the coat reading frame to +1 at the junction. These clones lyse quickly, as does pMS32.259 (Fig. 5). There are three more clones in which the coat reading frame is forced to +1 by the deletion. In two of these, pMS32.55 and pMS32.85, the onset of lysis was delayed with respect to the previous +1 frameshift mutants (Fig. 5), which may be related to deletions in the Shine-Dalgarno sequence, AGGU, around position 1,664 (Table 3). The last +1 frameshift clone is pMS32.88, which lacks the nonsense codon at position 1,672 together with the first 12 codons of the L gene, thus the absence of lysis was not unexpected. The coat protein ochre fragments or fusion peptides coded for by the respective plasmids can be detected in minicells (for examples see Fig. 6). In two clones, pMS32.17 and pMS32.50, the deletion leads to a -1 shift in the reading frame. In pMS32.17 this shift caused ribosomes to terminate at ochre stop codon 1,652. Evidently, termination at this position also allowed translation of the L gene mRNA, albeit with a lower frequency as judged by the delayed lysis of this clone (Fig. 5). Clone pMS32.50 did not lyse but in this case the -1 frame stop codon at position 1,652 had been deleted. It seems, therefore, that the frameshift itself was not signalling lysis protein expression but rather its inevitable consequence, chain termination.

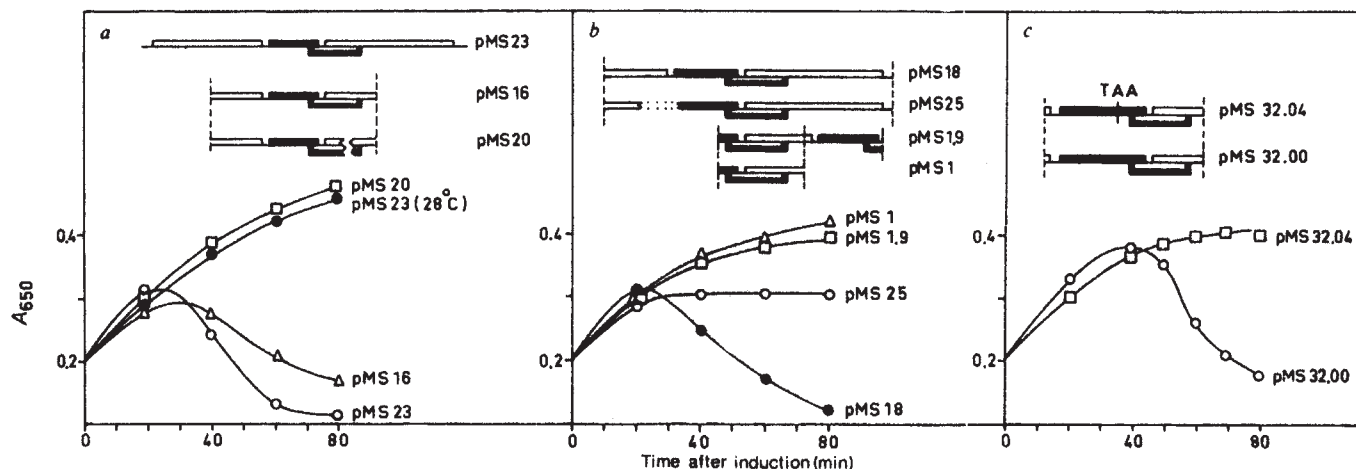


Fig. 3 Lytic properties of clones containing various regions of MS2 DNA inserted in pPLa831 or pPLa932 (see Table 1). Cells were grown at 28 °C to an A_{650} of 0.2. Growth was then continued at 42 °C (except where indicated). MS2 DNA regions present are shown schematically at the top.

Deletions in coat gene that do not change the reading frame

Two clones, pMS32.27 and pMS32.3, have deletions that do not change the reading frame. These two plasmids have lost the property to lyse their host (Fig. 5), which supports the view that expression of the L gene requires a ribosomal frameshift.

Obviously the absence of lysis in 0 frame coat mutants poses a problem as the wild-type sequence is read by definition in the zero frame. It seems, therefore, that the deletion decreases the chances that a phase shift occurs. The frameshift in wild-type MS2 RNA could arise in two ways: (1) interference of coat protein, bound to the start of the synthetase, with the movement of the reading ribosome. Although this involves an unknown mechanism it is formally correct, as the absence of wild-type coat protein distinguishes the 0 frame deletion mutants from the wild type. However, there is one argument that weakens this possibility. We have constructed a 0 frame mutant that carries a deletion of 18 nucleotides at the 5' end of the coat gene

(∇ 1,365–1,384), which is still able to lyse its host. (2) Frameshift is induced by a specific RNA structure present in or around the deleted region at position 1,628. An interesting sequence CAAUUUUC, present around position 1,620, or its immediate neighbourhood, was found to induce a +1 frameshift in yeast mitochondria¹⁴. However, the limited number of deletion mutants analysed does not allow a strong conclusion to be made concerning the significance of the 8-base homology. With regard to this, note that a +1 frameshift was observed *in vivo* between nucleotides 1,675 and 1,724, which does not include this particular nucleotide sequence⁵.

Translation of overlapping messages

As initiation of protein synthesis is slow compared with elongation, the independent translation of overlapping messages presents a general problem. Frequent readout of the upstream messenger (coat cistron) seems bound to interfere with independent initiation at the downstream message (L gene). At first glance one would therefore expect a regulatory mechanism

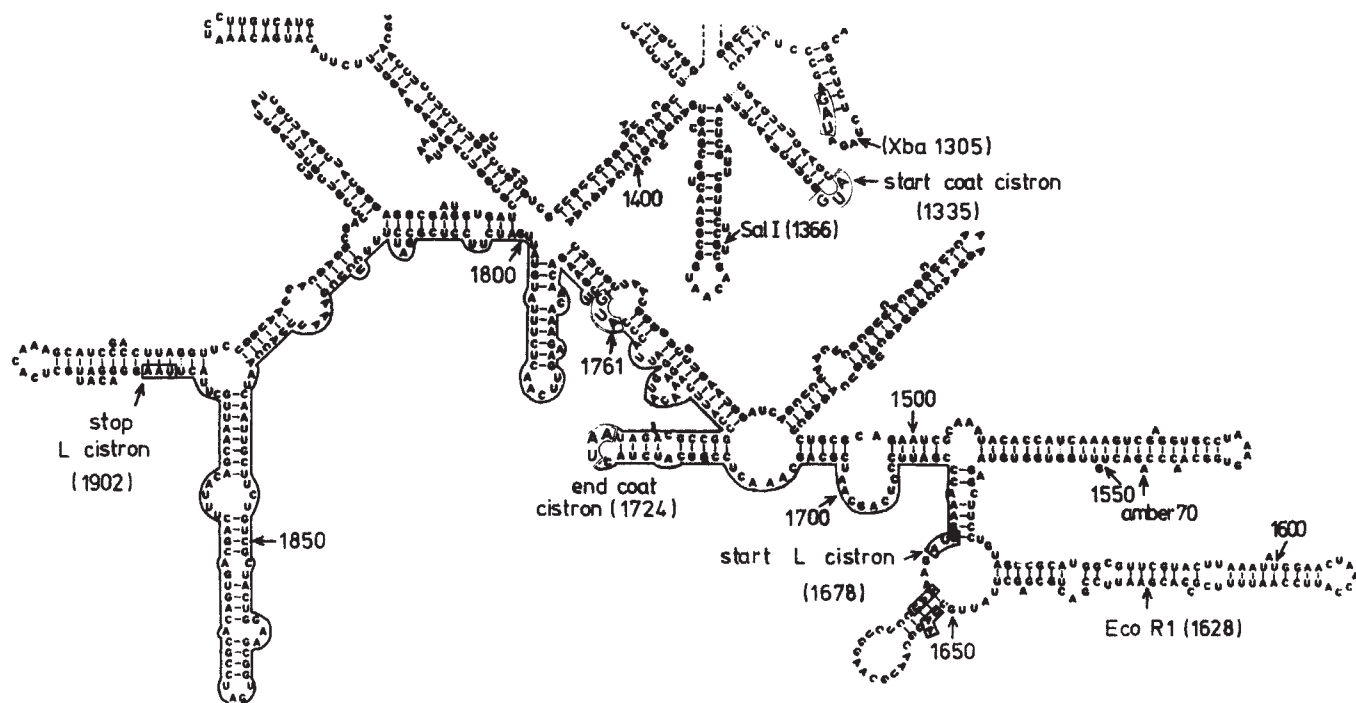


Fig. 4 Secondary structure model of the coat and lysis cistron of MS2 RNA⁶.

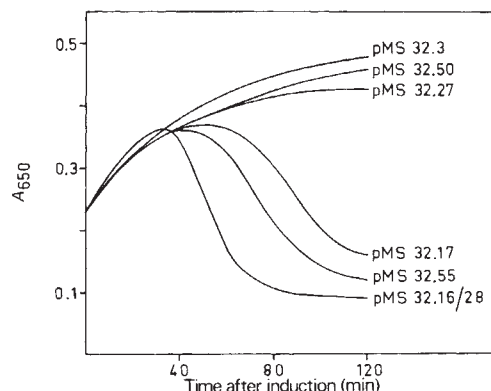


Fig. 5 Effect on cell lysis of reading frameshifts in the coat protein cistron. The MS2 DNA present in the plasmids is shown in Table 3. Growth of the cultures was as described in Fig. 3 legend.

that promotes expression of the downstream gene by interfering with the translational start at the upstream message to clear the template of ribosomes. However, such a solution seems impossible or inappropriate in this case. Instead, phage MS2 has solved the expression of the overlapping genes in an extraordinary way. The clones of the pMS32.N series lead us to conclude that translation of the L cistron relies on the fraction of the ribosomes that arrives out of frame at the beginning of this gene. The positions of two ochre codons, respectively in the +1 and -1 frame, just before the L cistron, elicits peptide chain termination, which in turn allows the reinitiation at the start of the L message, a region otherwise inaccessible to ribosomes. It is unlikely that ribosomes leave the RNA at this point as this would again create the problem of independent initiation.

The phenomenon of reinitiation has been recognized for some time in the *lacI*, *lacZ* and the bacteriophage T4 rIIB RNA. It may occur after a polypeptide chain is prematurely terminated at the site of a nonsense mutation¹⁵⁻¹⁸. Although several translational restarts have been analysed at the nucleotide level, it is not fully understood which features transform a sequence into a potential reinitiation site^{19,20}. The presence of a Shine-Dalgarno region seems dispensable for a translational restart²⁰. This agrees with the ability of our clones pMS32.55 and pMS32.85, which lack all or part of this region, to express the L gene, although expression is less efficient than when the Shine-Dalgarno sequence is still present (pMS32.16 and pMS32.28).

It is interesting to note that the phage may be able to adjust the level of L gene expression by varying the distance between

Table 2 Relationship between rate of MS2 coat protein synthesis and onset of cell lysis in four recombinants

Plasmid no.	MS2 coat protein present 30 min after induction at 42 °C (% of newly synthesized total protein)	Onset of lysis (min)
pMS2	30	30
pMS16	13	35
pMS12	6	40
pMS17	3	75

The relative amount of newly synthesized MS2 coat protein is calculated from protein separation on polyacrylamide gels (see Fig. 2). The coat protein is the major band in most clones and is identified by co-migration with *in vitro*-labelled MS2 coat protein extracted from the phage.

the AUG start codon and the preceding stop codon. This distance is 47, 23 and 3 nucleotides in pMS32.04, the -1 and the +1 frameshift mutants, respectively. The +1 frameshift mutant lyses quickly, the -1 slowly and pMS32.04 not at all. Although these represent only three cases, we suggest that the shorter the distance between termination and initiation codons, the higher the probability for reinitiation. The present data do not tell us if both or one particular phase shift operates *in vivo*. A recent analysis²¹ using *in vitro* translation of MS2 RNA seems to argue for a -1 phase shift in the region of interest. The occurrence of a frameshift polypeptide (peptide 9) terminating at position 1,652 has been reported elsewhere²¹.

Translational fidelity and expression of the lysis gene

The frameshift model presented here predicts that ribosomes which exhibit extreme fidelity in translation will not (or only barely) synthesize the L protein. This point is implicit in the recent observation of De Mars Cody and Conway²², that on infection of a particular streptomycin-resistant strain with wild-type f2 phage, virus particles accumulate inside the cell but no lysis follows. Ribosomes from streptomycin-resistant bacteria have a decreased error frequency, both with respect to mis-sense and frameshift²³.

Analogy with Φ X174 and G4?

Whether this model can be applied to other overlapping genes is uncertain. However, note a peculiar similarity between the overlap region in MS2 and that at genes A and B in the DNA

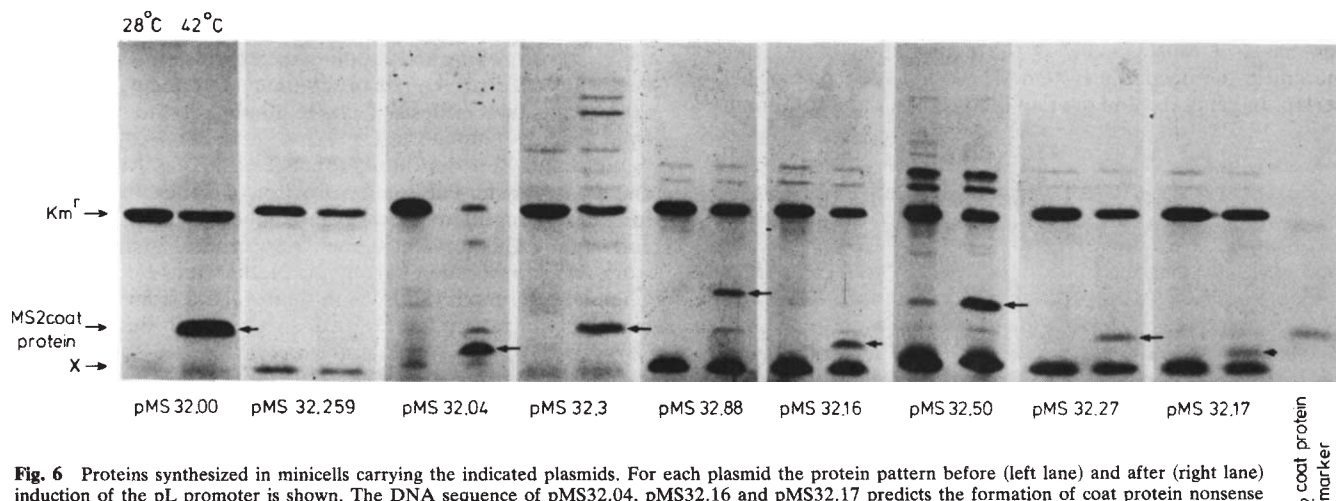


Fig. 6 Proteins synthesized in minicells carrying the indicated plasmids. For each plasmid the protein pattern before (left lane) and after (right lane) induction of the pL promoter is shown. The DNA sequence of pMS32.04, pMS32.16 and pMS32.17 predicts the formation of coat protein nonsense fragments of slightly different sizes. pMS32.50 and pMS32.88 give elongated proteins which, according to the DNA sequence, arise from fusion between the N-terminal region of the coat protein and a C-terminal part derived from a -1 frame (pMS32.50) or from the lysis protein (pMS32.88). Band X derives from a (stable) host mRNA. pMS32.3 and pMS32.27 are expected to yield MS2 coat protein lacking 1 and 9 amino acids, respectively. The protein formed in pMS32.3 cross-reacted with antiserum against MS2 coat protein. The lysis protein is not visible in any of the gels—this is possibly due to the very low level of synthesis of the protein. Translation from the start of the synthetase cistron produced a small amount of fusion protein running at the position of the coat protein (see for example pMS32.88). Arrows facing left indicate the MS2 coat protein or nonsense fragments, and the fusion proteins derived from it.

Table 3 Effect on cell lysis of deletions in the coat protein sequence around nucleotide 1,628

	1,610	1,620	1,630	1,640	1,650	1,660	1,670	1,680	
	CAUUCCAAUUUUCGCCACGAAUUCGACUGCGAGCUUAUUGU <u>UAA</u> CGCAAUGCAAGGUCUCCUAAAGAUG								
					Stop		Stop	Start lysis gene	
					(-1 frame)		(+1 frame)		
No. of nucleotides deleted	Reading frame								Cell lysis
3	CAUUCCAAUUUUCGCC 0		AAUUCGACUGCGAGCUUAUUGUAAAGGCAAUGCAAGGUCUCCUAAAGAUG						-
16	CAUUCCAAUUUUCGCC +1		AGCUUAUUGUAAAGGCAAUGCAAGGUCUCCUAAAGAUG						+
17	CAUUCCAAUUUUCGCC -1		AGCUUAUUGUAAAGGCAAUGCAAGGUCUCCUAAAGAUG						+
27	CAUUCCAAUUUUCGCC 0		AAGGCAAUGCAAGGUCUCCUAAAGAUG						-
28	CAUUCCAAUUUUCGCC +1		AGGCAAUGCAAGGUCUCCUAAAGAUG						+
50	CAUUC		GUCUCCUAAAGAUG						-
55	CAUUC		CUAAAGAUG						+
85			+ Deleted up to nucleotide 1,580 GUCUCCUAAAGAUG						+
88	CAUUCCAAUUUUCGCC +1		Deleted up to nucleotide 1,714 +						-

pMS32.3 was constructed by opening the unique *EcoRI* site of pMS32.00 at position 1,628. The linearized DNA was treated with T4 DNA polymerase and re-sealed with T4 DNA ligase. The other plasmids were constructed by treatment of pMS32.00 with the enzyme *Bal31* as described in the text. All sequences were determined by the procedure described elsewhere²⁵. 0, +1 and -1 in the deleted sequence refer to the change in reading frame at the junction. Stop codons are boxed. In pMS32.N, N indicates the number of nucleotides deleted. The position of the deleted nucleotides is shown.

phages Φ X174 and G4 (Fig. 7). Strikingly, the nucleotide indicated by an arrow in the out-of-phase ochre codon of Φ X174 is the third base of a codon specifying alanine, thus any other base would do. The corresponding U in G4 specifies asparagine, thus a C here would not change the meaning of the codon. Statistically, the chance of finding a U in the marked position in both phages is thus 1 in 8, if the only purpose is to specify the particular amino acids. This suggests that the U is there to introduce a stop codon.

Why gene overlap?

The design adopted by the phage to control and express the L gene has two aspects. One is the exploitation of the noise inherent in the decoding system. The other is the use of gene overlap. In general, gene overlap is considered as a solution for

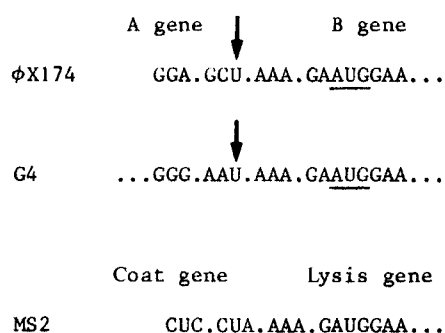


Fig. 7 Nucleotide sequences at the A/B gene overlap of phage Φ X174 and G4. The marked U is nucleotide 5,058 of Φ X174 (ref. 26) and 1,270 of the G4 sequence (ref. 27). Dots indicate the reading frame of the A gene. The coat-lysis overlap of MS2 is given for comparison.

storing as much information as possible in a limited amount of RNA or DNA. It seems that another aspect of the overlap may be equally important: it allows and enforces the extreme coupling of expression of two genes that may be difficult or impossible to realize were the genes separated. For example, the chosen design guarantees that there will not be lysis of the host unless enough coat protein has been made, that is, until the infection has proven successful²⁴. The use of the phase shift siphons off a very small fraction of the ribosomes destined for L protein synthesis. This, together with the position(s) of the nonsense codon(s), sets the balance between the amount of coat and L proteins needed.

The exploitation of reading frame errors represents a novel means of controlling and coupling the expression of overlapping genes. If such errors can be programmed into the base sequence this permits the synthesis of a large number of different proteins from a single template.

We thank Ben Berkhout, Johan Twilt, Kees Ravensbergen and Koos Van der Plas for construction of clones, Rein Dijkema for introducing us to the Maxam-Gilbert sequence technique and Hanny Bakker-Steeneveld for technical assistance. This work was sponsored in part by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO).

Received 15 July; accepted 13 October 1981.

1. Horiuchi, K. in *RNA Phages* (ed. Zinder, N. D.) 29-50 (Cold Spring Harbor Monogr. Ser., 1975).
2. Jeppesen, P. G. N., Argetsinger-Steitz, J., Gesteland, R. F. & Spahr, P. F. *Nature* **226**, 230-237 (1970).
3. Model, P., Webster, R. E. & Zinder, N. D. *Cell* **18**, 235-246 (1979).
4. Atkins, J. F., Steitz, J. A., Anderson, C. W. & Model, P. *Cell* **18**, 247-256 (1979).
5. Beremand, M. N. & Blumenthal, T. *Cell* **18**, 257-266 (1979).
6. Fiers, W. *et al. Nature* **260**, 500-507 (1976).
7. Robertson, H. D. in *RNA Phages* (ed. Zinder, N. D.) 113-145 (Cold Spring Harbor Monogr. Ser., 1975).
8. Zinder, N. D. & Lyons, L. B. *Science* **159**, 84-86 (1968).

9. Devos, R., Van Emmelo, J., Contreras, R. & Fiers, W. *J. molec. Biol.* **128**, 595–619 (1979).
10. Remaut, E., Stanssens, P. & Fiers, W. *Gene* **15**, 81–93 (1981).
11. Steitz, J. A. in *RNA Phages* (ed. Zinder, N. D.) 319–352 (Cold Spring Harbor Monogr. Ser., 1975).
12. Lodish, H. F. in *RNA Phages* (ed. Zinder, N. D.) 301–318 (Cold Spring Harbor Monogr. Ser., 1975).
13. Min Jou, W., Haegeman, G., Ysebaert, M. & Fiers, W. *Nature* **237**, 82–88 (1972).
14. Fox, T. D. & Weis-Brummer, B. *Nature* **288**, 60–63 (1980).
15. Sarabhai, A. & Brenner, S. *J. molec. Biol.* **27**, 145–162 (1967).
16. Files, J., Weber, K. & Miller, J. *Proc. natn. Acad. Sci. U.S.A.* **71**, 667–670 (1974).
17. Michaels, C. & Zipser, D. *J. molec. Biol.* **41**, 341–347 (1969).

18. Manley, J. L. *J. molec. Biol.* **125**, 449–466 (1978).
19. Steege, D. A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4163–4167 (1977).
20. Napoli, C., Gold, L. & Swebelius Singer, B. *J. molec. Biol.* **149**, 433–449 (1981).
21. Atkins, J. F., Gesteland, R. F., Reid, B. R. & Anderson, C. W. *Cell* **18**, 1119–1131 (1979).
22. De Mars Cody, J. & Conway, T. W. *J. Virol.* **37**, 813–820 (1981).
23. Gorini, L. in *Ribosomes* (eds. Nomura, M., Tissières, A. & Lengyel, P.) 791–803 (Cold Spring Harbor Monogr. Ser., 1974).
24. Sugiyama, T., Stone, H. & Nakada, D. *J. molec. Biol.* **42**, 97 (1969).
25. Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
26. Sander, F. *et al. Nature* **265**, 687–695 (1977).
27. Godson, G. N., Barrel, B. G., Staden, R. & Fiddes, J. *Nature* **276**, 236–247 (1978).

LETTERS TO NATURE

Particle acceleration at planetary bow shock waves

M. M. Hoppe & C. T. Russell

Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024, USA

We can extend our understanding of collisionless shocks by comparing their behaviour in a variety of plasma conditions at several different planets. One property of such shocks is the occurrence of upstream magnetohydrodynamic waves associated with particle beams accelerated at these shocks, and flowing back towards the Sun¹. We report here observations of one of these classes of wave at Mercury, Venus, Earth and Jupiter. First, we show that the empirical relationship between the interplanetary field strength and the frequency of the waves in the observer's rest frame, derived originally from ground-based pulsation studies, and later, *in situ*, is approximately true at all the planets. Second, the observed frequencies are consistent with resonance with beams of ions of essentially the same energy at each of the planets. This observation is consistent with Sonnerup's² geometrical model of ion reflection at collisionless shocks. The universality of this ion acceleration mechanism in the Solar System suggests that like acceleration occurs in similar astrophysical systems outside our Solar System and thus may provide a source of acceleration for cosmic rays.

One characteristic feature of the solar wind interaction with the planets is the creation of an upstream bow shock, formed when the supersonic solar wind encounters the planetary atmosphere or magnetic field. Such shocks have been seen at all the planets so far probed by interplanetary spacecraft, and their general character is similar in many ways to that of Earth's bow shock. However, certain facets of shock structure *per se*, such as the associated plasma wave spectra³ and the magnetic field overshoot⁴, do vary between planets. These changes are generally consistent with the evolution of solar wind parameters with heliocentric distance.

We present a complementary comparative study of planetary foreshocks, those regions upstream of, but magnetically connected to, the bow shocks. The varieties of reflected particles and associated waves characteristic of this region at the Earth have been studied in some detail⁵. We focus here on the large-amplitude ($\Delta B/B \sim 1$), low-frequency (~ 0.03 Hz) waves and the associated backstreaming ion populations. We now briefly review results of the terrestrial work which are pertinent to the current study and compare them with studies of the upstream regions of Mercury, Venus and Jupiter.

We use primarily magnetic field data because the plasma analysers onboard extraterrestrial missions tend to be configured to measure the solar wind and are not in general appropriate instruments for characterizing the lower-density reflected particles. We focus on the large-amplitude low-frequency waves which have been documented in the upstream regions of Mercury, Venus, Earth and Jupiter, and studied in depth at the Earth^{6,7}. It seems likely that these waves are generated through cyclotron resonance with beams of reflected ions⁸, although theory does not yet reproduce all the features of

the observations⁹. These waves have been of interest for some time, partially because of their role in the foreshock structure, but also because of their apparent connection with geomagnetic pulsations of the same frequency.

The relationships between upstream conditions and pulsation character have been investigated on the hypothesis that the upstream waves may be the ultimate source of at least some classes of pulsations. A linear relationship between pulsation frequency and the magnitude of interplanetary magnetic field has been found^{10,11}. Russell and Hoppe¹² recently verified that the same relationship that holds for ground data also holds for the upstream waves. We have found that this same relationship is maintained to a remarkable degree over the orders of magnitude differences in wave frequency and field strength observed during the various planetary encounters. This is shown in Fig. 1 where the frequencies of the upstream waves observed at the different planets are plotted as a function of the associated magnetic field strengths. Data for Mercury represent Fairfield and Behannon's¹³ estimate of 5–10 s periods for the 'low-frequency' waves encountered by Mariner 10 on its outbound pass. The Venus data were obtained from the Pioneer Venus magnetometer. The Earth data are from ref. 12 and the two points from Jupiter represent intervals of upstream wave activity, one observed by Pioneer 10 (ref. 14), the other by Voyager 2. The straight line is the empirical relationship between wave frequency and field magnitude found at Earth¹². Points for all the planets clearly fall along the same linear trend.

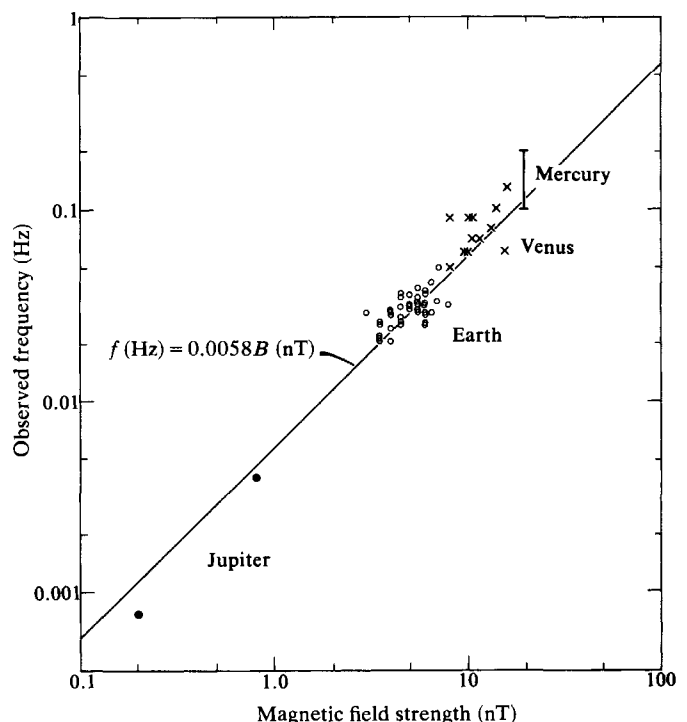


Fig. 1 Upstream wave frequency as observed in the spacecraft frame (f_{obs}) of reference as a function of ambient magnetic field strength.

Table 1 Observed and rest-frame frequencies and resonant velocities

Planet	f' (Hz)	f_0/f_{ci}	v_r/v_{sw}
Mercury	0.1–0.2	0.09–0.17	1.2–2.2
Venus	0.06–0.13	0.064, 0.074	1.7, 1.9
Earth	0.02–0.04	0.03–0.08	2.5 ± 0.3
Jupiter	0.0008(P) 0.003(V)	0.06 0.02	2.3 2.1

Plasma data for Mercury from ref. 21, for Venus from Barnes (personal communication), for Earth from the Los Alamos ISEE team (personal communication) and for the Voyager Jupiter encounter (V) from the MIT plasma group (personal communication). Typical terrestrial values were scaled to provide estimates for the Pioneer Jupiter encounter (P).

This relationship can be understood as a combined effect of ion resonance and Doppler shifting, assuming that the waves are in each case low-frequency magnetosonic waves similar to those observed at the Earth. The terrestrial waves, which are a characteristic feature of the foreshock region, have long been thought to result from instability of ion beams reflected back along the IMF by the bow shock^{8,15}, although resonance with relativistic electrons has also been recently suggested as a source of the jovian waves¹⁴. The terrestrial waves are, however, very closely linked with upstream ion distributions¹⁶ and spacecraft studies¹⁷ have demonstrated that their properties are generally consistent with an ion beam generation mechanism. Specifically, it was shown that the terrestrial upstream waves are indeed magnetosonic waves with rest frame frequencies (f_0) of the order of 0.1 the local proton cyclotron frequency (f_{ci}). Because of Doppler shift effects this identification required the use of two point measurements, which for the terrestrial case were provided by the dual ISEE 1 and 2 spacecraft. One cannot ordinarily determine a wave's rest frame frequency using data from a single satellite. This is because even given the observed frequency (f'), solar wind velocity (v_{sw}) and wave propagation direction $\hat{k}$, there remain two unknowns, the rest frame frequency (f_0) and wavelength (λ) in the Doppler shift equation:

$$f' = f_0 + \frac{\mathbf{v}_{sw} \cdot \hat{\mathbf{k}}}{\lambda} \quad (1)$$

The number of unknowns can be reduced if we assume that the upstream waves at the extraterrestrial planets have rest frame frequencies of the same order as the terrestrial waves, $f_0 \approx 0.1 f_{ci}$. Waves in these low-frequency ranges have phase velocities very close to the Alfvén speed, $v_A = B^2/(4\pi n)^{1/2}$, where B is the magnetic field strength and n is the plasma density. Using this fact allows us to substitute $f_0 v_A = \lambda$ for λ equation (1). We can solve for f_0 in terms of the measured parameters f' , v_{sw} , k , B and n :

$$f_0 = f' / \left(1 + \frac{\mathbf{v}_{sw} \cdot \hat{\mathbf{k}}}{B/(4\pi n)^{1/2}} \right) \quad (2)$$

Rest frame frequencies were estimated using this approach for all observations for which solar wind data were available, and wave propagation directions were determined through minimum variance analysis when possible. The 180° ambiguity which results in minimum variance analysis may be resolved by assuming the waves at Mercury, Venus and Jupiter are attempting to propagate upstream, as has been demonstrated to be the case at Earth. The angle between $\mathbf{v}_{sw}$ and $\hat{\mathbf{k}}$ was determined by minimum variance analysis of the Mercury, Earth and Voyager Jupiter data, was estimated from the average field orientation for Pioneer 10, and was assumed to have the same average value (35°) at Venus. Table 1 summarizes rest frame frequencies estimated in this way, normalized to the local proton cyclotron frequency. Estimated rest frame frequencies are of the same order at each planet, averaging about $0.07 f_{ci}$.

Having found results consistent with the hypothesis that the waves are in each case low-frequency magnetosonic waves, we can estimate the proton beam velocities required to generate waves of the observed frequencies. This was done by substituting calculated rest frame frequencies into the formula for the resonant velocity (v_r) of protons interacting with a magnetosonic mode wave¹⁸:

$$\frac{v_r}{V_A} = \frac{(1 + f_0/f_{ci})^{3/2}}{f_0/f_{ci}} \quad (3)$$

Calculated resonant velocities normalized by the ambient solar wind speeds are also presented in Table 1. Note that the resonant velocity given is the velocity parallel to the magnetic field relative to the component of the solar wind velocity along the field whereas most measurements are of total energy of the particles in the shock frame of reference. The beam energies are remarkably similar in all four cases, suggesting that essentially the same mechanism operates at each bow shock.

The consistency of the relationship between the observed wave frequencies and the magnetic field strength at the various planets derives from this equivalence of the source beam energies. In fact, taking a value of 2 for v_r/v_{sw} , solving for f_0 using equation (3) and substituting this value of f_0 into equation (2) using a nominal $\cos \theta_{kx} = -0.8$ yields the relationship f' (Hz) = $0.006 B$ (nT), which is in excellent agreement with the empirically determined relationship. Furthermore, the similarity of the inferred beam energies argues that the magnitude of bow shock acceleration must be relatively constant despite differences in shock size and structure. This similarity is consistent with Sonnerup's description of bow shock acceleration through particle displacement along the interplanetary electric field during the reflection process. In this description, which appears to predict accurately reflected beam energies at Earth¹⁹, shock structure and strength have no influence and the amount of ion energization is determined strictly by geometrical effects. On the other hand, it is intriguing to see the smallest energies associated with the small relatively weak bow shock of Venus²⁰. In any case, whatever the particle acceleration mechanism, the analysis presented here suggests that ion acceleration is a characteristic feature of planetary bow shocks and furthermore that reflected particle energies are of the same order regardless of shock scale size or structure.

Mariner 10 magnetic field data for Mercury and Voyager magnetic field data for Jupiter were obtained from NSSDC. This work was supported by NASA through contracts NAS 5-25772 and NGL 5-007-190.

Received 28 July; accepted 3 November 1981.

- Hoppe, M. M. & Russell, C. T. *Adv. Space Res.* **1**, 327–332 (1981).
- Sonnerup, B. U. Ö. *J. geophys. Res.* **74**, 1301–1304 (1969).
- Scarf, F. L., Gurnett, D. A. & Kurth, W. S. *Nature* **292**, 747–750 (1981).
- Russell, C. T., Hoppe, M. M. & Livesey, W. A. *Nature* (in the press).
- J. geophys. Res.* **86**, 479–4536 (1981).
- Fairfield, D. H. *J. geophys. Res.* **74**, 3541–3553 (1969).
- Hoppe, M. M., Russell, C. T., Frank, L. A., Eastman, T. E. & Greenstadt, E. W. *J. geophys. Res.* **86**, 4471–4492 (1981).
- Barnes, A. *Cosmic Electrodyn.* **1**, 90–114 (1970).
- Gary, S. P., Gosling, J. T. & Forslund, D. W. *J. geophys. Res.* **86**, 6691–6696 (1981).
- Troitskaya, V. A., Plyasova-Bakovina, T. A. & Gul'yel'mi, A. V. *Dokl. Akad. Nauk. SSSR* **197**, 1312 (1971).
- Gul'yel'mi, A. V. & Bol'shakova, O. V. *Geomagn. Aeron.* **13**, 535 (1973).
- Russell, C. T. & Hoppe, M. M. *Geophys. Res. Lett.* **8**, 615–618 (1981).
- Fairfield, D. H. & Behannon, K. W. *J. geophys. Res.* **81**, 3897–3906 (1976).
- Smith, E. J., Tsurutani, B. T., Chenette, D. L., Conlon, T. F. & Simpson, J. A. *J. geophys. Res.* **81**, 65–72 (1976).
- Gul'yel'mi, A. V. *Space Sci. Rev.* **16**, 331–345 (1974).
- Paschmann, G. et al. *Geophys. Res. Lett.* **6**, 209–210 (1979).
- Hoppe, M. M. & Russell, C. T. *EOS* **62**, 363 (1981).
- Kennel, C. F. & Petschek, H. E. *J. geophys. Res.* **71**, 1–28 (1966).
- Paschmann, G., Schopke, N., Asbridge, J. R., Bame, S. J. & Gosling, J. T. *J. geophys. Res.* **85**, 4689–4694 (1980).
- Russell, C. T., Elphic, R. C. & Slavin, J. A. *Nature* **282**, 815–816 (1979).
- Ogilvie, K. W. et al. *Science* **185**, 145–151 (1974).

Organic and inorganic interpretations of the martian UV-IR reflectance spectrum

Kevin D. Pang*, Sandy F. S. Chun†, Joseph M. Ajello*, Zhao Nansheng‡, & Liang Minji§

* Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

† School of Medicine, University of California, San Francisco, California 94143, USA

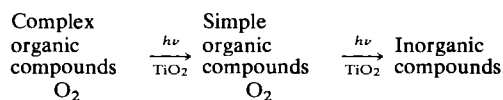
‡ Beijing Planetarium, China

§ Beijing Glass Institute, China

Although some data from the Viking Lander biology experiments can be interpreted as indicative of biological activity, the existence of organisms in the martian soil samples is considered unlikely because of the non-detection of organic compounds in the sample¹. Viking gas chromatography-mass spectrometer analysis detected no organic molecules above a concentration of parts per 10⁹ (ref. 2). We consider here why no organic molecules were detected at the landing sites, whether the sterility of the two sites is representative of the entire planet and if there are locations on Mars more conducive to the formation and preservation of organics. We first simulate the destruction of organic compounds in Mars-like laboratory conditions, and then examine whether the destructive mechanism would operate planetwide; and second re-examine UV and IR reflectance spectra of Mars for any evidence of organic molecules, and in doing so set an upper limit on the organic carbon content of average martian soil. The results reveal that the average martian soil is organic-poor and makes an unfavourable habitat for life forms based on carbon chemistry. There is no reason to believe that organic molecules are preferentially preserved anywhere on the planet.

Recent observations of the outer Solar System by ground-based and spacecraft UV-IR spectroscopy show an abundance of organic compounds, existing as atmospheric gases and imbedded in carbonaceous chondrites. Because of its proximity to the outer Solar System, Mars should have received a large infall of carbonaceous meteorites^{3,4}. The fact that Viking found no organics at the landing sites means that a very efficient process must be operating on Mars to destroy the organic molecules.

Chun *et al.*^{3,4} proposed a mechanism for the destruction of organics on Mars: UV-stimulated catalytic oxidation. The degradation process from complex organic ultimately into inorganic compounds is:



In a controlled laboratory experiment, powders of Murchison meteorite (C2), rich in organics, and TiO₂ were degassed in Pyrex flasks. The flasks were filled with oxygen at approximately martian surface partial pressure, 0.1 torr. These flasks were then irradiated with UV-visible light. Hard UV light that may cause photolysis was first filtered out. The flasks were rotated on a turning wheel to keep the powder constantly agitated, and inhomogeneities in the radiation field averaged out. The flasks were examined for changes in gaseous composition using a mass spectrometer. Samples of the powder were pyrolysed and analysed with a mass spectrometer. According to our proposed mechanism, an evolution of CO₂ in the flasks, and/or a change in distribution of masses from high to low molecular weights in the

solids would mean that UV-stimulated catalytic oxidation of organic compounds has taken place.

After irradiation for 4 days a copious supply of CO₂ was observed in flasks where all three ingredients, O₂, TiO₂ and carbonaceous chondrite, were present. The evolution of CO₂ was accompanied by a decrease in O₂ abundance. In flasks where all ingredients, except TiO₂ were present a small but detectable reaction took place. In flasks where all ingredients except O₂ were present, or where the carbonaceous chondrite was replaced by basalt, there was no evolution of CO₂. Also no reaction was observed in flasks with all ingredients present but kept in the dark.

We interpret the results of the controlled experiment to mean that organic compounds in the carbonaceous chondrite are photocatalytically oxidized. Activation energy, provided by energetic photons, is required for the reaction to take place. A catalyst, TiO₂, speeds up the reaction, but is not essential.

Mass spectra of pyrolysed solid samples were difficult to interpret due to the vast number of peaks present. Attempts to identify and trace the breakdown of high-molecular-weight compounds, such as undecane (abundant in C2), into low-molecular-weight compounds, such as acetone (a product generally observed from photocatalytic oxidation of alkanes), were not successful due to obscuration by fragments from the breakdown of many other organic compounds. Instead of trying to trace the breakdown of a mixture of organic compounds present in the carbonaceous chondrite, we decided to trace the breakdown of one compound at a time. An example is the breakdown of naphthalene, an aromatic compound. The naphthalene sample was prepared and irradiated along with O₂ and TiO₂ as its carbonaceous chondrite counterpart. The evolution of CO₂ was again observed, accompanied by the appearance of

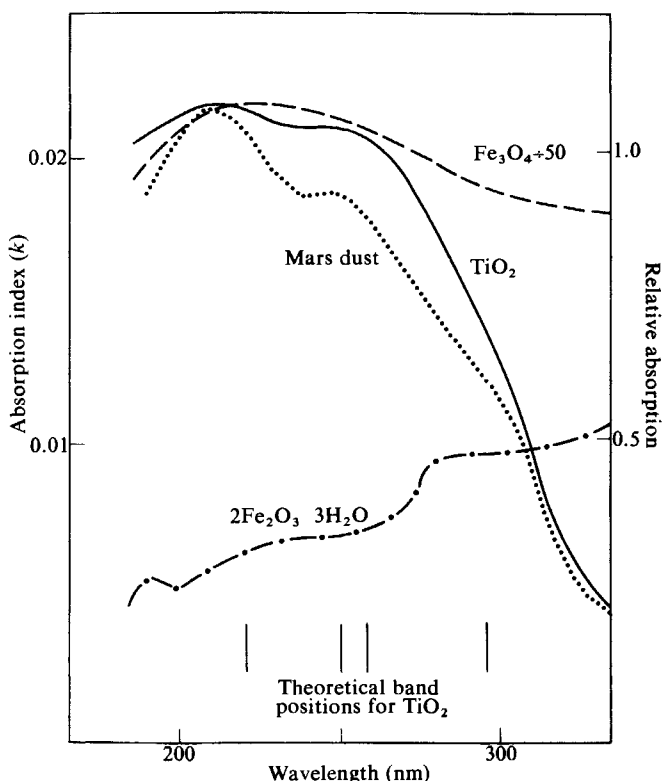


Fig. 1 Dotted curve indicates the absorption index *k* of martian dust (ref. 13). Solid curve is the relative absorption of anatase (an allotrope of TiO₂), normalized to the peak of Mars dust (measured by Zhao and Liang at the Beijing Glass Institute). Molecular orbital calculations for TiO₂ predict bands at the wavelengths indicated by tick marks (ref. 21). Data for magnetite (dashed curve) and limonite (dot-dashed curve) are from refs 23 and 24, respectively.

simple organic compounds such as methane. We conclude that the naphthalene has broken down into simpler compounds.

Recently Sancier and Wise⁵ found that photocatalytic oxidation is responsible for the degradation of organic compounds in terrestrial sand dunes. The catalyst for the sunlight-assisted reaction can be ordinary sand particles.

Whereas all conditions required for the photocatalytic oxidation of organic compounds: O₂ (ref. 6), sand (the dunes in Viking Lander pictures are evidence), and UV irradiation (as indicated by the darkening of UV-sensitive chips mounted on the Viking Landers, ref. 7) are present at the Viking landing sites, we conclude that the proposed mechanism operates at least at these two locations. To extrapolate this finding to the rest of the planet we need to examine conditions elsewhere to see if the requirements for photocatalytic oxidation are also satisfied.

The polar regions of Mars have been suggested as the most favourable sites for preservation of organics but there is no basis for this speculation. First, Mariner 9 UV spectrometer observations showed that solar radiation of wavelengths as short as 200 nm easily penetrates the martian atmosphere down to the polar cap⁸. Second, sand grains exist in great abundance in the polar regions on Mars—in dunes and laminated between condensates⁹. Third, ozone is found in the martian polar atmosphere, although not in a concentration to attenuate the solar UV significantly¹⁰. The presence of O₃ requires that O₂ be there also. Therefore, all conditions necessary for the photocatalytic oxidation of organics—UV radiation, oxygen, and sand—are present in the polar regions of Mars, and the destruction of any organic molecules on the polar caps seems almost certain.

Theoretical atmospheric modelling by Hunten¹¹ showed that the formation of oxidants is planetwide on Mars. Atmospheric radiative transfer calculations by Kuhn and Atreya¹² found that all latitudes on Mars receive lethal annual dosage of solar UV radiation. Sand dunes are common on Mars⁹. Therefore, all conditions required for photocatalytic oxidation of organics seem to be satisfied, and we conclude that there is no basis for optimism for finding organic molecules anywhere on Mars. Organics could be buried underground, but the fact that Viking found none under a rock does not support this conjecture.

In addition to the above comparative analysis we can use spectral remote sensing data of Mars to extend the Viking findings. UV and IR spectroscopy are very sensitive techniques in quantitative organic analysis. Therefore, UV and IR reflectance spectra of Mars, when properly interpreted with the aid of laboratory 'ground truth' measurements, can yield the surface chemical composition of areas not examined *in situ*.

Most molecules have spectral signatures in the small range of UV wavelengths 120–320 nm. Unfortunately UV observations of the martian surface are usually hampered by Rayleigh scattering (at $\lambda \leq 320$ nm) and absorption (at $\lambda \leq 180$ nm) due to carbon dioxide molecules in the martian atmosphere^{10,13}. However, during dust storms surface material is raised and suspended as dust in the atmosphere. The 1971 global dust storm was particularly favourable for remotely sensing the physical properties of Mars dust particles, and the UV spec-

trometer on the Mars-orbiting Mariner 9 spacecraft observed the dust clouds as resolutions and phase angles not possible from the Earth. By modelling the observed phase function (intensity as a function of the scattering angle) of the dust particles with theoretical scattering calculations their complex refractive index ($n - ik$) can be determined. The wavelength dependence of the absorption index k of martian dust is shown in Fig. 1. The absorption increases with decreasing wavelength down to ~ 210 nm, but decreases at still shorter wavelengths. A structural shoulder in the absorption spectrum occurs near 250 nm. A third feature is also present at ~ 300 nm. Both the absolute value of k and its variation in the UV are consistent with k values in the visible/IR, determined by Viking Lander spectrophotometric observations of martian aerosols¹⁴. UV observations of the dust-shrouded Mars by the Orbiting Astronomical Observatory (OAO 2) are also consistent with the Mariner 9 measurements¹⁵.

Abadi and Wickramasinghe¹⁶ have suggested that the absorption feature at 210 nm may be a signature of highly complex organic or pre-biotic molecules in martian dust clouds. Their conclusion is based on similarities between the 210 nm martian dust band and bands in the absorption curves of organic materials, including an extract of the Murchison meteorite^{17,18}. Although there is a superficial resemblance between the martian dust absorption spectrum and the data in refs 17 and 18, comparisons reveal important differences in both band strength and position. Searching through more than 2,000 UV absorption spectra of organic compounds¹⁹ we could not find any that closely resemble the dotted curve in Fig. 1. A combination of organic compounds or a sample that we did not examine may have had the required spectral characteristics. However, the following inorganic interpretation is more reasonable because of its simplicity, and because it accounts for all the UV features observed in martian dust.

The suggestion that the UV absorption bands in martian dust are due to electronic transitions in titanium dioxide¹³ has been experimentally verified at the Beijing Glass Institute. Titanium oxide was vacuum evaporated onto a fused quartz substrate, and then oxidized by heat treatment²⁰. X-ray diffraction study revealed the crystalline structure of the TiO₂ to be in the anatase form. The optical density of the sample was measured with a spectrophotometer at a resolution of $\Delta\lambda/\lambda = 0.001$. The solid curve in Fig. 1 shows the relative absorptivity of the anatase sample, normalized to the peak of martian dust absorption. Molecular orbital calculations predict absorption bands at 220, 250, 258, and 296 nm due to metal–ligand transitions in TiO₂ (tick marks in Fig. 1)²¹. We should also consider whether other electronic transitions (Fe²⁺, Fe³⁺) that have been suggested to explain the visible/near IR absorption features of Mars²² may also account for the UV absorption features. Absorption spectra of Fe₃O₄ and 2Fe₂O₃ · 3H₂O are also shown in Fig. 1. The ferric and ferrous transitions do not fit the Mars dust data as well as TiO₂. In any case, TiO₂ is expected to dominate the UV spectrum of Mars surface material despite its lower abundance, because its band strength is far stronger than that of the iron oxides below about 320 nm (refs 23–25). Our conclusion is

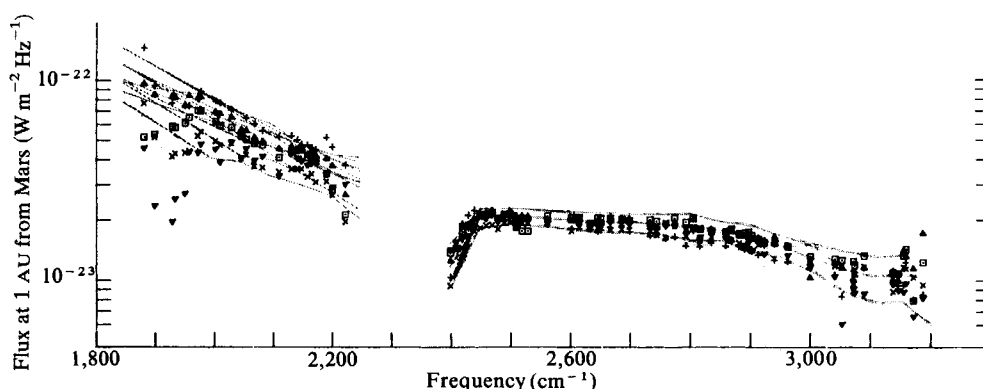


Fig. 2 Observed IR flux of Mars compared with inverse wavelength. Different symbols refer to different observing runs. Different curves were generated by a temperature-albedo model. Reproduced from refs 27, 28, with permission of Reidel, Boston.

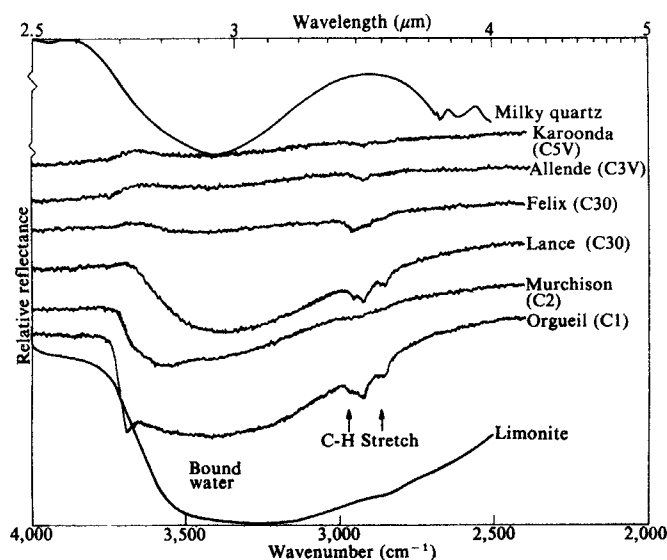


Fig. 3 Relative reflectance of powders (particle sizes between 0 and 74 μm) of quartz, limonite, and carbonaceous chondrites. Petrographical classification in parentheses. Reflectance (relative to MgO) at λ 2.5 μm are: Murchison 7%, Orgueil 9%, Allende 11%, Karoonda 14%, Felix 15%, Lance 17%, limonite 34%, and quartz 82% (see ref. 46). Reflectance measurements of carbonaceous chondrites made by Salisbury are published in part in ref. 39. Quartz and limonite spectra are given in ref. 40.

consistent with the detection of titanium by Viking Lander inorganic analysis²⁶.

In addition to the UV analysis we may reinterpret IR reflectance spectra of Mars to derive useful upper limits on the organic carbon content of martian soil. Extensive IR spectroscopic studies of Mars have been carried out with sophisticated groundbased, airborne, and spacecraft instrumentation. The martian IR spectrum shows the signatures of atmospheric carbon dioxide, carbon monoxide, water vapour, and oxygen; and of silicates, solid CO_2 , bound and frozen water on the surface^{22,27-32}. Stringent upper limits have been placed on undetected atmospheric constituents, for example, parts per 10^8 for methane³³⁻³⁶. The wavelength interval 3.3–3.6 μm has been especially well scrutinized because nearly all organic compounds containing C and H have absorption bands here due to excitation of the C–H stretching vibration^{37,38}. As all gases known to be present in the martian atmosphere⁶ are transparent in this spectral region, any absorption feature found superimposed on the wing of the 3- μm water of hydration band may be assigned to organic constituents in Mars surface material. Despite intensive search the C–H absorption bands have never been found on Mars. Figure 2 shows high resolution spectra of Mars taken by Beer²⁷. Absorption features due atmospheric CO_2 and water bound in surface material ($3,300\text{ cm}^{-1}$ or 3 μm) are clear. However, the C–H bands ($2,800\text{--}3,000\text{ cm}^{-1}$ or 3.3–3.6 μm) are absent. Observed flux variations in this spectral interval can be adequately accounted for by a simple temperature–albedo model.

Laboratory measurements are useful for quantitative reinterpretations of IR reflectance spectra of Mars. Figure 3 shows laboratory reflectance spectra of quartz, limonite, and representative carbonaceous meteorites^{39,40}. The methyl and methylene (symmetric and anti-symmetric) vibration bands are clearly resolved as a doublet on the long wavelength wing of the water of hydration band in the spectra of the carbonaceous chondrites. The strength of the C–H bands correlates with the organic carbon content of the meteorite, for example, Orgueil, 2.5%; Lance and Felix 0.4%; and Allende, 0.2%. The C–H bands are detectable in even in Karoonda, with an organic carbon content of only 0.08%. The 3.3–3.6 μm features are not present in quartz and limonite, representing terrestrial minerals

and rocks with organic carbon contents of 0.02–0.06% (ref. 41, 42). Organic carbon content has been estimated to be about 80% of total carbon content by Nagy. The remaining 20% carbon is in carbonates which do not have the C–H bond. Total organic carbon contents of terrestrial rocks are estimated from vaporization-pyrolysis measurements of representative rocks, scaled to that of the Allende meteorite⁴³. We conclude that the absence of the fundamental C–H stretching absorption features in the martian reflectance spectra indicates that the organic carbon content of its average surface material is at or below parts per 10^4 . The upper limit for organic carbon content pertains to a value averaged over the whole martian planet, as when Mars is observed by low spatial resolution Earth-based IR spectroscopy, or to a value averaged over the field of view (FOV) of spacecraft IR spectrometers, which have best resolution of the order of 100 km. The organic carbon content of an area smaller than the resolution limit might exceed that averaged over the FOV, although Viking GC–MS results make this proposition seem unlikely. Field observations in the dry valleys of Antarctica have shown that soil samples with organic carbon contents <0.1% are generally sterile^{44,45}. This implies that Mars in general and areas examined by IR spectroscopy in particular have soil deficient in organic matter, and are unfavourable habitats for life forms based on carbon chemistry.

An inorganic interpretation of Mars' UV reflectance spectrum is, therefore, more plausible than an organic interpretation. The IR reflectance spectrum places a stringent upper limit on the organic carbon content of average martian soil, virtually ruling out biological activities. Laboratory simulation and site comparisons give no reason to believe that organic molecules are preferentially preserved anywhere on Mars. Taken together, the evidence suggests that the sterility of the Viking landing sites is representative of the entire planet.

We thank Drs William H. Pickering and Ta-Jin Kuo for initiating this collaborative effort, and Drs Jack W. Salisbury, Eddie N. Wells, Larry A. Lebofsky, and James L. Gooding for helpful comments. K.D.P. performed part of the research during his tenure as a National Academy of Sciences Senior Resident Research Associate. S.F.S.C.'s participation was made possible through funding from the 1977 Newcomb Cleveland Award, given by the American Association for the Advancement of Science. This research was performed in the People's Republic of China and at JPL. The work carried out at JPL, California Institute of Technology, was sponsored by NASA. The Murchison meteorite specimen used in our research was provided by the Smithsonian Institution.

Received 21 September; accepted 30 October 1981.

- Klein, H. P. *Rev. Geophys. Space Phys.* **17**, 1655–1662 (1979).
- Biemann, K. & Lavoie, J. M. Jr *J. geophys. Res.* **84**, 8385–8390 (1979).
- Chun, S. F. S., Pang, K. D., Cutts, J. A. & Ajello, J. M. *Nature* **274**, 875–876 (1978).
- Chun, S. F. S., Pang, K. D. & Ajello, J. M. *NASA Conf. Publ.* **2072**, 15 (1977).
- Sancier, K. M. & Wise, H. *Atmos. Envir.* **15**, 639–640 (1981).
- Owen, T. et al. *J. geophys. Res.* **82**, 4635–4639 (1977).
- Arvidson, R. & Jones, K. *NASA Conf. Publ.* **2072**, 3 (1979).
- Pang, K. & Hord, C. W. *Icarus* **18**, 481–488 (1973).
- Carr, M. H. *Space Sci. Rev.* **25**, 231–284 (1980).
- Barth, C. A. & Hord, C. W. *Science* **173**, 197–201 (1971).
- Hunten, D. M. *J. molec. Evol.* **14**, 71–78 (1979).
- Kuhn, W. R. & Atreya, S. K. *J. molec. Evol.* **14**, 57–64 (1979).
- Pang, K. & Ajello, J. M. *Icarus* **30**, 63–74 (1977).
- Pollack, J. B. et al. *J. geophys. Res.* **84**, 2929–2946 (1979).
- Caldwell, J. *Icarus* **32**, 190–209 (1977).
- Abadi, H. & Wickramasinghe, N. C. *Nature* **267**, 687–688 (1977).
- Sakata, A. et al. *Nature* **266**, 241 (1977).
- Hoyle, F. & Wickramasinghe, N. C. *Nature* **266**, 241–243 (1977).
- Scott, A. I. *Interpretation of the Ultraviolet Spectra of Natural Products* (Pergamon, Oxford, 1964).
- Hass, G. *Vacuum* **2**, 331–345 (1952).
- Loeffler, B. M., Burns, R. G., Tossell, J. A., Vaughan, D. J. & Johnson, K. H. *Proc. 5th Lunar Sci. Conf.* 3007–3016 (1974).
- Singer, R. B., McCord, T. B., Clark, R. N., Adams, J. B. & Huguénin, R. L. *J. geophys. Res.* **84**, 8415–8426 (1979).
- Huffman, D. K. & Stapp, J. L. in *Interstellar Dust and Related Topics* (eds Greenberg, M. & van de Hulst, H. C.) 297–301 (Reidel, Boston, 1973).
- Egan, W. G., Hilgeman, T. & Pang, K. D. *Icarus* **25**, 344–355 (1975).
- Cardona, M. & Harbecke, G. *Phys. Rev.* **137**, A1467–A1476 (1965).
- Toulmin, P. III et al. *J. geophys. Res.* **82**, 4625–4634 (1977).
- Beer, R. in *Exploration of the Solar System* (eds Woszczyk, A. & Iwanaszewska, C.) 229–239 (Reidel, Boston, 1974).

28. Barker, E. S. in *Exploration of the Solar System* (eds Woszczyk, A. & Iwanowska, C.) 229–239 (Reidel, Boston, 1974).
29. Noxon, J. F., Traub, W. A., Carleton, N. P. & Connes, P. *Astrophys. J.* **207**, 1025–1035 (1976).
30. Egan, W. G., Hilgeman, T. & Smith, L. L. *Icarus* **35**, 209–226 (1978).
31. Pimentel, G. C., Forney, P. B. & Herr, K. C. *J. geophys. Res.* **79**, 1623–1634 (1974).
32. Hanel, R. et al. *Icarus* **17**, 423–442 (1972).
33. Maguire, W. C. *Icarus* **32**, 85–97 (1977).
34. Beer, R., Norton, R. H. & Martonchik, J. V. *Icarus* **15**, 1–10 (1971).
35. Horn, D., McAfee, J. M., Winer, A. M., Herr, K. C. & Pimentel, G. C. *Icarus* **16**, 543–556 (1972).
36. Owen, T. & Sagan, C. *Icarus* **16**, 557–568 (1972).
37. Duley, W. W. & Williams, D. A. *Nature* **277**, 40–41 (1979).
38. Rea, D. G., O'Leary, B. T. & Sinton, W. M. *Science* **147**, 1286–1288 (1965).
39. Lebofsky, L. A. *Mon. Not. R. astr. Soc.* **182**, 17P–21P (1978).
40. Salisbury, J. W. & Hunt, G. R. *Nature* **222**, 132–136 (1969).
41. Nagy, B. *Carbonaceous Meteorites* (Elsevier, Oxford, 1975).
42. Wiik, H. B. *Commentat. physico-math.* **34**, 135–145 (1969).
43. Levy, R. M., Grayson, M. A. & Wolf, C. J. *Geochim. cosmochim. Acta* **37**, 467–483 (1973).
44. Cameron, R. E. in *Research in the Antarctica* (eds Quam, L. O. & Porter, H. D.) 137–189 (American Association for the Advancement of Science, Washington DC, 1971).
45. Horowitz, N. H., Cameron, R. E. & Hubbard, J. S. *Science* **176**, 242–245 (1972).

Chorus-related electrostatic bursts in the Earth's outer magnetosphere

L. A. Reinleitner, D. A. Gurnett & D. L. Gallagher

Department of Physics and Astronomy, University of Iowa,
Iowa City, Iowa 52242, USA

We report here recent studies of wideband plasma wave data from the ISEE 1 and ISEE 2 spacecraft which reveal that whistler-mode chorus emissions in the Earth's outer magnetosphere are often accompanied by high-frequency bursts of electrostatic noise. The chorus features that correlate to the electrostatic bursts have a hook-like frequency-time variation in the frequency range 200–400 Hz, and the associated electrostatic bursts have a frequency range of ~4–8 kHz. In some cases these electrostatic bursts have a harmonic frequency structure with a frequency spacing corresponding to the chorus frequency. This harmonic structure apparently results from the fact that the burst intensity is modulated at the chorus frequency.

Selected wideband spectrograms¹ obtained from ISEE 1 of the chorus-related electrostatic bursts are shown in Figs 1–3. Figure 1 indicates some of the general features of the bursts. In this case, the bursts appear to be strongly correlated to discrete hook-like features in the chorus band². In addition to this example, many cases of the electrostatic bursts are observed in conjunction with intense chorus bands without hooks. In such cases the electrostatic bursts are usually merged together in a nearly continuous band similar to the extended burst in Fig. 1 from about 1738:40 to 1738:50 UT. A comprehensive survey of the locations where the bursts (with or without hooks) are observed has not yet been performed. However, all cases observed have been in the outer magnetosphere near the day-side magnetopause boundary from 7.0 to 17.5 magnetic local time (MLT), about 9–12 R_E and within ~30° of the magnetic equator.

The frequent occurrences of the electrostatic bursts, whether accompanied by hooks in the chorus band or not, as well as the

rather limited physical region of occurrence, strongly indicate that the bursts are not caused by instrumental effects. To check that the bursts are not due to some cross-modulation effect in the instrument electronics, wideband data from the IMP 6 spacecraft have been examined. (The plasma wave instrumentation on IMP 6 is detailed in refs 3 and 4.) Although IMP 6 was not in the correct wideband mode to observe both frequencies of interest as often as ISEE 1, intense chorus accompanied by electrostatic bursts have, nevertheless, been found in several passes through the outer magnetosphere on the dayside region.

The electrostatic bursts have several characteristics. The bursts typically occur in a frequency band extending from near the electron plasma frequency f_p as determined by the lower edge of the continuum radiation⁴, down to approximately half f_p , as in Fig. 1. In the examples shown, the frequency range of the burst does not extend up into the continuum radiation, although bursts extending up into the plasma frequency have been found. In all cases, the bursts occur at frequencies above the electron gyrofrequency ($f_g \sim 1.2$ kHz in Fig. 1).

The bursts have been described as electrostatic because no wave magnetic field can be detected in association with them. Typically the broadband electric field strength of the bursts are about $50 \mu\text{V m}^{-1}$. Because the wave magnetic field remains at the instrument noise level in all cases, a limit can be placed on the electric-to-magnetic field ratio of about $E/cB \geq 0.25$ (which corresponds to an index of refraction $n \leq 4$). Because no electromagnetic plasma wave mode is known to exist with $E/cB \geq 0.25$ in the frequency range $f_g < f_{\text{burst}} < f_p$, the bursts are almost certainly electrostatic. The typical field strengths of the chorus emissions associated with the electrostatic bursts are about $300 \mu\text{V m}^{-1}$ and 40×10^{-12} T.

The electrostatic bursts occasionally have narrowband harmonic structure of the type shown in Fig. 2. The harmonic structure should not be confused with the vertical striations due to the rapid time dependence of the bursts. The harmonic structure is only evident when the chorus emission has a narrow bandwidth, such as when the hook-shaped features occur. The frequency spacing of the harmonic structure corresponds exactly with the instantaneous emission frequency of the chorus burst.

When chorus hooks are found to be correlated with the electrostatic bursts, the onset of the electrostatic burst usually coincides with the minimum frequency of the hook. This relationship can be seen most clearly in the first and last hook in the upper panels of Fig. 3. Note that the electrostatic burst usually continues for a time after the hook has disappeared or merged back into the chorus band.

Comparative studies using both ISEE 1 and ISEE 2, were possible because both spacecraft were in nearly identical orbits with a separation of only hundreds of kilometres. Although virtually identical chorus hooks were found in the electric field wideband data, the electrostatic bursts, while correlated, were not identical at the two spacecraft which indicates that the structure of the electrostatic bursts varies significantly on spatial scales of only hundreds of kilometres. To estimate the wavelength of the electrostatic noise, the electric field strength has been compared on ISEE 1 and 2, which have electric antenna lengths of 215 m and 30 m tip-to-tip. For bursts which had similar characteristics at the two spacecraft, the electric field

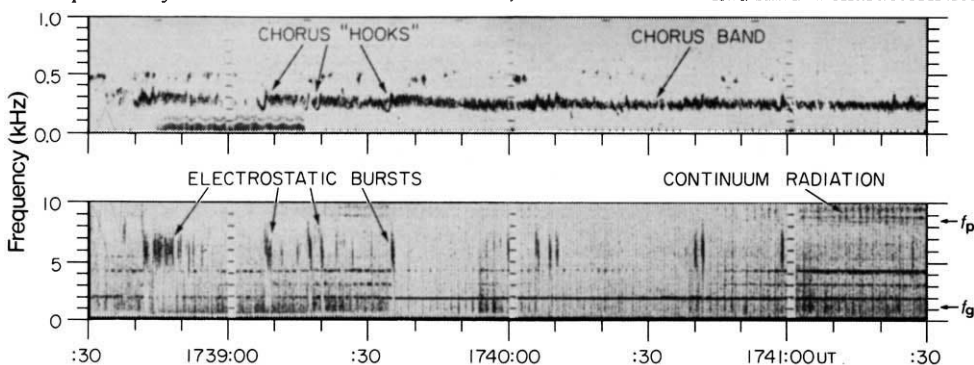


Fig. 1 Frequency versus time wideband plasma wave data for 5 November 1977 showing general features of the chorus related electrostatic bursts. The slight interference features of the chorus related electrostatic bursts. The slight interference from 1738:45 to 1739:17 UT (also in Fig. 3) is from the ISEE electron density experiment. $R = 11.04 R_E$; magnetic lat. = 21.1°; MLT = 11.6 h.

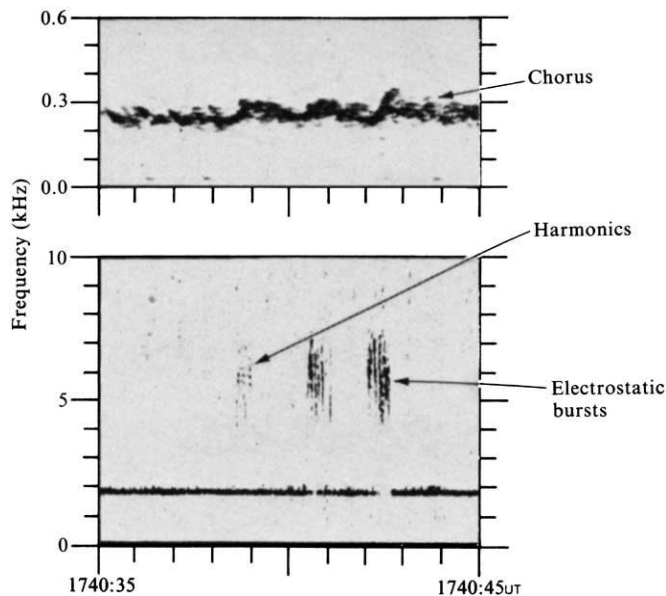


Fig. 2 Wideband plasma wave data showing narrowband harmonic structure for 5 November 1977. $R = 11.01 R_E$; magnetic lat. = 21.1° ; MLT = 11.7 h.

strengths were in good agreement, thereby indicating that the wavelength of the bursts must be longer than both of the antennas, implying a wavelength for the electrostatic bursts of $\lambda_{\text{burst}} > 215$ m.

Comparisons of the waveforms of the chorus and the electrostatic bursts reveal several interesting features. The lower panel of Fig. 3 shows the electric field versus time for both the chorus and the electrostatic burst. The lower signal is that of the chorus while the upper high frequency signal is the associated electrostatic burst. Both signals have been run through bandpass

filters to eliminate signals not in the frequency band of interest. Note that the time duration of these waveform data is only 8 ms. The waveform therefore comprises only a small portion of the electrostatic burst from which it was taken. As can be seen, the electrostatic noise is actually composed of many shorter bursts of a high frequency signal which is modulated at the chorus frequency. Thus, the harmonics evident in Fig. 2 are a modulation effect caused by the periodic pulsing of the electrostatic noise at the frequency of the chorus emission. In the cases where the harmonics do not occur, the chorus does not occur as a single frequency but rather as a band of frequencies, which tends to blend harmonics into a broad continuous spectrum.

The correlation between the envelope of the bursts and the chorus waveform is often not as good as the example shown; however, the basic modulation effect in which the intensity of the high frequency emission maximizes at a given phase of the chorus waveform is usually evident. Also, despite the apparently abrupt initial turn-on of the bursts in the usual wideband data, waveform data show a gradual increase in the amplitude of the high frequency bursts, as well as a gradual decrease toward the end of each burst. For the case shown in the lower panel of Fig. 3, the frequency of the electrostatic emission is 6.7 kHz (± 0.5 kHz) which is near but below the plasma frequency of 8.5 kHz (± 0.5 kHz) as determined by the lower cutoff of the continuum radiation.

The strong correlation between the envelope of the electrostatic bursts and the chorus waveform suggests some strong physical interaction between these two wave modes. We currently believe that the chorus is phase bunching electrons at the phase velocity of the whistler mode and that these resonant electrons are triggering an electrostatic instability through a two-stream (or bump-on-tail) type mechanism⁵. We are not certain of the exact identification of the electrostatic mode involved but the best candidate seems to be a Langmuir wave or related beam mode, because the emission occurs very close to the local electron plasma frequency. The shift in the frequency of the electrostatic emission below the plasma frequency could

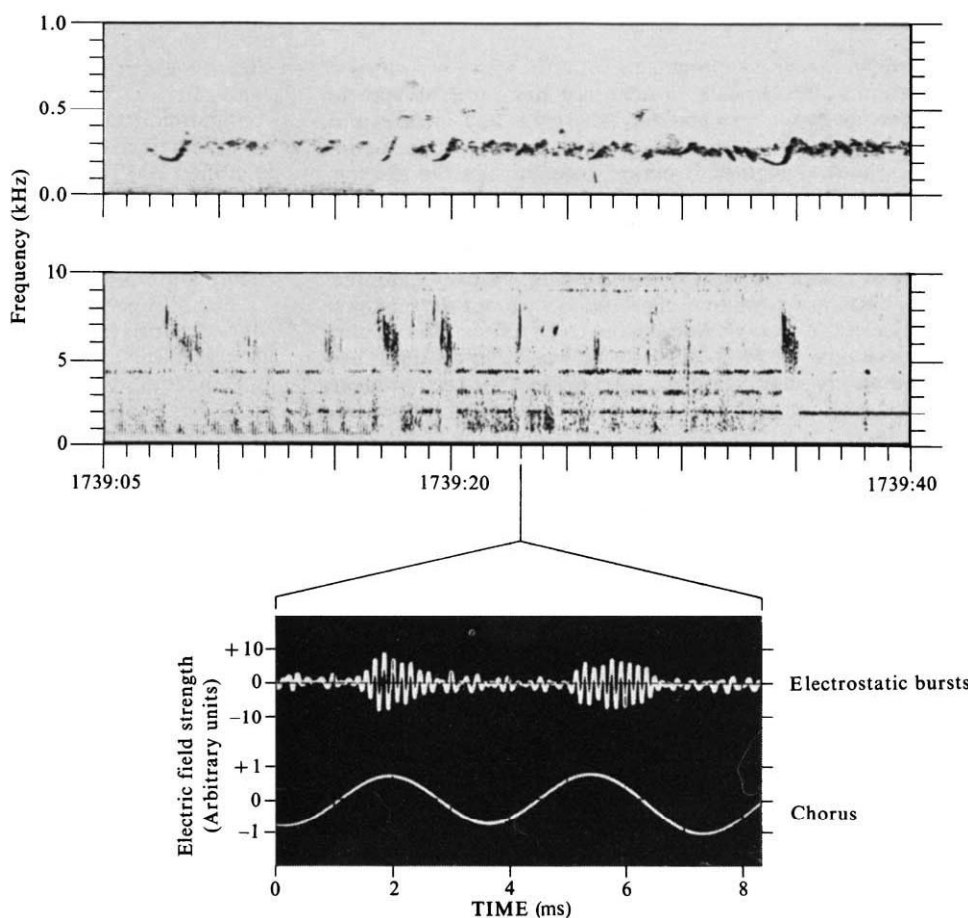


Fig. 3 The lower panel shows the oscilloscope waveform pattern taken from the very short burst indicated in the upper panels of wideband data for 5 November 1977. The low frequency signal in the bottom of the lower panel is the chorus waveform, while the high frequency signal bursts at the top of the lower panel are the electrostatic bursts. The phase of the chorus waveform may differ from that shown by a constant (instrumental) phase factor. $R = 11.04 R_E$; magnetic lat. = 21.1° ; MLT = 11.6 h.

be due to either a Doppler shift or an intrinsic effect related to the two-stream interaction.

We also believe that the correlated whistler and electron plasma oscillation bursts detected in the solar wind by Kennel *et al.*,⁶ are probably essentially identical to the chorus related electrostatic bursts described here. Most differences are attributable to differing plasma parameters in the two different regions of interest. The hypothesized interaction between these two wave modes is seen as a better alternative than the secondary impulsive electron heating mechanism proposed by Kennel *et al.*⁶.

We thank Dr Alain Roux for valuable suggestions on theory, and Dr Roger R. Anderson for help with analysis of the data. This research was supported by NASA through grant NGL-16-001-043 from NASA Headquarters and contract NAS5-20093 with Goddard Space Flight Center, and the Office of Naval Research.

Received 3 July; accepted 16 October 1981.

1. Gurnett, D. A., Scarf, F. L., Fredricks, R. W. & Smith, E. J. *IEEE Trans. Geosci. Electron.* **GE-16**, 225-230 (1978).
2. Helliwell, R. A. *Whistlers and Related Ionospheric Phenomena* (Stanford University Press, 1965).
3. Shaw, R. R. & Gurnett, D. A. *J. geophys. Res.* **80**, 4259-4271 (1975).
4. Gurnett, D. A. & Shaw, R. R. *J. geophys. Res.* **78**, 8136-8149 (1973).
5. Krall, N. A. & Trivelpiece, A. W. *Principles of Plasma Physics* (McGraw-Hill, New York, 1973).
6. Kennel, C. F. *et al. Geophys. Res. Lett.* **7**, 129-132, (1980).

Coordination model for the defect structure of hyperstoichiometric UO_{2+x} and U_4O_9

G. C. Allen, P. A. Tempest & J. W. Tyler

Central Electricity Generating Board, Berkeley Nuclear Laboratories, Berkeley, Gloucestershire GL13 9PB, UK

Non-stoichiometric compounds or solids in which the ratios of the amounts of elements present are not integral may be characterized by the formula MX_x where x is non-integral and, in many crystalline compounds, may vary over quite a wide range without change of structure type. Ideally the thermodynamic stability and physical properties of such compounds should be related to the same atomic and electronic model but the ways in which various simple structures adapt to departures from ideal composition have not been adequately explored. Models based on a random distribution of point defects have proved successful but selected area electron diffraction and ultra high resolution electron microscopy have shown that their distribution is not simple. Interaction between defects produces microscopic sub-structures or clusters introducing specific local properties. Although our knowledge of the principal morphological features of sub-structures has advanced, their precise interatomic arrangements are generally unknown. Here we derive a model for non-stoichiometric oxides of uranium.

Some oxides of the lanthanide or actinide elements have fluorite-type structures and a propensity towards deviation from simple stoichiometry. Oxygen can be transferred easily between the MO_x lattice and the ambient gas phase. UO_2 is no exception and its stoichiometric form is not easy to obtain. Although some oxidation occurs even at very low temperatures, homogeneous oxidation products are not formed in this way. At room temperature, surface oxidation processes are responsible for the non-stoichiometric nature of UO_2 in dry air and further oxidation is controlled by diffusion of oxygen either in the dioxide lattice or through a diphasic surface layer. Only above $\sim 300^\circ\text{C}$ does the oxidation become so rapid that uniform accessibility of oxygen to the entire sample becomes a factor. In many defect oxides with the fluorite structure non-stoichiometry occurs

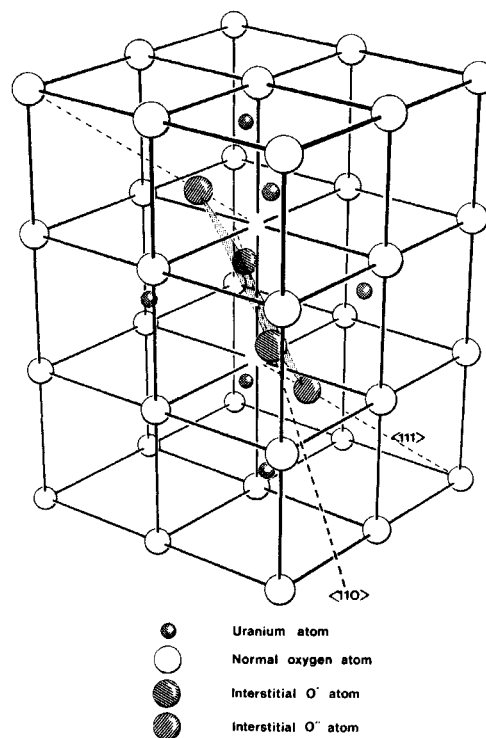


Fig. 1 The oxygen sublattice showing a 2:2:2 cluster.

through untenanted anion sites, the cation sublattice remaining essentially intact. The formation of UO_{2+x} and U_4O_9 , however, occurs by the incorporation of additional oxygen atoms at interstitial sites in the fluorite structure of UO_2 rather than the creation of uranium vacancies, and the occupation number of cations again remains the same as in the stoichiometric compound.

The interstitial sites have been identified by Willis^{1,2}. These are of two types which are displaced from the cubic-coordinated sites by about $1 \text{ \AA}$ in the $\langle 110 \rangle$ and $\langle 111 \rangle$ directions. This aggregate or 'cluster' depicted in Fig. 1 has the 2:2:2 configuration which contains two oxygen interstitial atoms along $\langle 110 \rangle$ identified as O' , two oxygen interstitial atoms along $\langle 111 \rangle$ identified as O'' , and two vacancies in the oxygen sublattice. The formation of these clusters contrasts with the structural defects present in other anion-excess fluorite-related superlattices in which the basic oxygen cube of the fluorite structure changes to a square antiprism³.

Martin⁴ suggested that ordering of defects in fluorite lattices could occur to form chains of coordination defects. For MO_{2-x} phases Martin's recognition of this phenomenon was based on assemblies of vacancy point defects. However, a detailed analysis of the X-ray photoelectron spectrum of UO_{2+x} suggested a defect grouping based on a unique coordination of Willis type clusters in which an individual oxygen atom has a total negative charge less than in the isolated oxygen ion⁵. Thus ordering similar to that suggested by Martin could occur in UO_{2+x} based on oxygen interstitials. Our model for the structure of UO_{2+x} including U_4O_9 , incorporates a chain-like coordination of oxygen defect clusters.

In Fig. 1 the 2:2:2 cluster bridges two uranium atoms which then become 9-fold rather than 8-fold coordinate. Using neutron diffraction data⁶ for this UO_9 coordination polyhedron the internuclear separations are; $6(\text{U}-\text{O})$ at $2.37 \text{ \AA}$, $2(\text{U}-\text{O}')$ at $2.22 \text{ \AA}$, and $1(\text{U}-\text{O}'')$ at $2.29 \text{ \AA}$ emphasizing the higher degree of covalency in the uranium-cluster bonds. These calculations require that the U-U distance be fixed although in practice, this is unlikely. Single crystal X-ray diffraction studies of defect fluorite-type structures reveal consistent lengthening of U-U and shortening of O-O distances around a vacant oxygen site⁷ and similar polarization effects would be expected in the vicinity of a cluster. This would shorten the U-U interatomic distance

across the cluster and create preferred interstitial oxygen sites along the opposite edge of the oxygen cube containing the original defect grouping. Thus in addition to the tendency for clusters to order under coulombic forces to develop a superstructure, on geometrical grounds we can postulate the formation of a 2:2:2 cluster chain along $\langle 110 \rangle$. A $\langle 110 \rangle$ chain in the oxygen sublattice is shown in Fig. 2 and its relationship to the conventional UO_2 unit cell is given in Fig. 3.

Within the chain each uranium atom now has the coordination number 10. Most models for the anion excess phase invoke doubly charged interstitials (O_i^{2-}) compensated by singly charged holes (h^+) present as U^{5+} cations². At first sight bond length data seem to support this view. The average U–O separation in most metal uranate (VI) structures is 2.07–2.10 Å but in the eight coordinate CaUO_4 structure this distance is 2.21 Å (ref. 8) which approaches the values derived above. Data from measurements of the X-ray photoelectron spectra in the U4f region may also be interpreted in terms of a uranium atom in crystallographically inequivalent sites where it apparently resides in 4+ and 6+ ionic configurations^{9,10} so that although ESR measurements on uranates favour the presence of U^{5+} in U^{6+} lattices¹¹ this may not be the case in the fluorite structure of UO_2 .

The length and frequency of chains determines the overall nature of the structure. For low values of x in UO_{2+x} a random distribution of clusters is envisaged. As x increases the probability of oxygen interstitials being influenced by clusters already formed increases. Gradually as x approaches 0.25 cluster chains develop and become ordered. Continuous chains are only allowed in every alternate uranium sublattice layer parallel to $\langle 001 \rangle$ because of O'' interference in the oxygen sublattice. At a distance of closest approach of $2\sqrt{2}a_0$ (where a_0 is the lattice parameter of the parent UO_2 fluorite lattice) in the $\langle 110 \rangle$ direction only 32 clusters can be accommodated by a unit cell of edge $4a_0$ which yields the composition of U_4O_9 exactly, and is consistent with the superlattice observed by Willis⁶. A chain spacing of $\frac{3}{2}\sqrt{2}a_0$ parallel to $\langle 110 \rangle$ yields U_3O_7 but is not consistent with the U_4O_9 superlattice. Any increase in oxygen content above $\text{UO}_{2.33}$ would involve severe distortion of the parent fluorite lattice as additional oxygen atoms would have to be accommodated by oxygen cubes already distorted by O' or O'' interstitials. Not surprisingly the next characterized oxide phase U_3O_8 has a completely different structure. The increased electrostatic force between uranium atoms separated by a 2:2:2 cluster produces a lattice shrinkage along the unique $[110]$ chain axis. Conversely O' interstitials along $[1\bar{1}0]$ produce a lattice expansion. Hence in U_4O_9 the presence of one or several preferentially oriented chains will distort the superlattice from

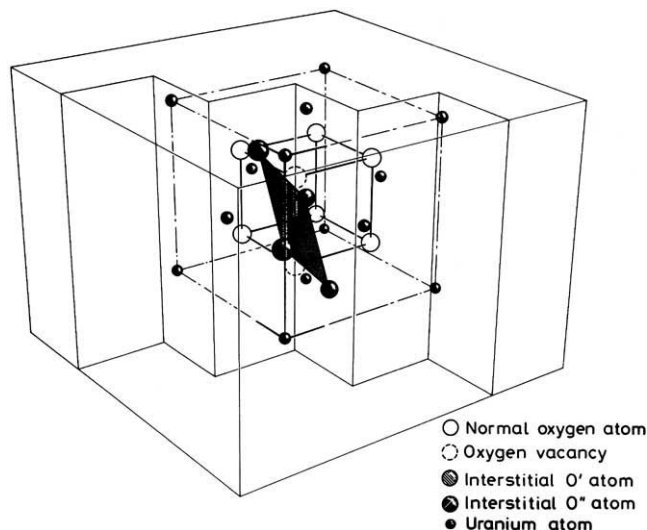


Fig. 3 The relationship between a 2:2:2 cluster and the normal fluorite unit cell.

cubic symmetry. Indeed distortion from the cubic to rhombohedral structure in the superlattice of U_4O_9 has been observed¹².

Although the U_3O_7 composition could be accommodated by the above model no complete experimental determination of its crystal structure exists. Lattice dimensions have been reported in terms of a face-centred tetragonal cell derived from the UO_2 structure but the tetragonal phase may be metastable.

This paper is published by permission of the Central Electricity Generating Board.

Received 3 September; accepted 13 November 1981.

1. Willis, B. T. M. *Nature* **197**, 755–756 (1963).
2. Willis, B. T. M. *Acta crystallogr.* **A34**, 88–90 (1978).
3. Bevan, D. J. M., Greis, O. & Strähle, J. *Acta crystallogr.* **A36**, 889–890 (1980).
4. Martin, R. L. *J. Proc. R. Soc. N.S.W.* **109**, 137–150 (1976).
5. Allen, G. C., Tucker, P. M. & Tyler, J. W. *JCS chem. Commun.* 691–692 (1981).
6. Willis, B. T. M. *J. Phys. Radiat.* **25**, 431–439 (1964).
7. Thornber, M. R. & Bevan, D. J. M. *J. Solid-St. Chem.* **1**, 536–544 (1970).
8. Allen, G. C., Griffiths, A. J. & Lee, B. J. *Transition Metal Chem.* **3**, 229–233 (1978).
9. Verbit, J., Riga, J., Pireaux, J. J. & Caudano, R. *J. Electron Spectros. relat. Phenom.* **5**, 193–205 (1974).
10. Allen, G. C., Tucker, P. M. & Tyler, J. W. *J. phys. Chem.* (in the press).
11. Allen, G. C., Griffiths, A. J. & Suckling, C. W. *Chem. Phys. Lett.* **53**, 309–312 (1978).
12. Belbeoch, B., Boivineau, J. C. & Perio, P. *J. Phys. Chem. Solids* **28**, 1267–1275 (1967).

Synergistic tellurium–caesium embrittlement of Type 316 stainless steel

M. G. Adamson

General Electric Company, Vallecitos Road, Pleasanton, California 94566, USA

E. A. Aitken & S. Vaidyanathan

General Electric Company, 310 DeGuigne Drive, Sunnyvale, California 94086, USA

Austenitic stainless steels have become important high-temperature structural and containment alloys in applications ranging from industrial processes to energy generation and conversion. The AISI Type-300 series of iron–base austenitic stainless steels, in particular, is widely used as a barrier or primary containment material for both nuclear and fossil fuels operation at temperatures as high as 700 and 1,000 °C, respectively. Thus any observed deleterious effect of chemical environment on the alloy mechanical properties such as the ‘hot’ embrittlement of austenitic stainless steel by liquid zinc^{1,2} is very

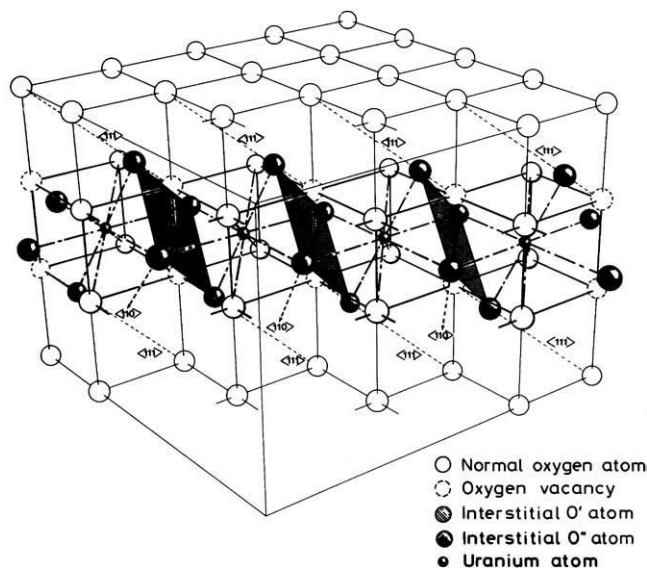


Fig. 2 A chain of 2:2:2 clusters in the oxygen sublattice.

important and, indeed, Old³ has mentioned that fission products such as caesium, cadmium and tellurium might embrittle the austenitic stainless steel cladding of fast reactor fuel pins in certain conditions. The few investigations of the high temperature mechanical behaviour of iron-base austenitic stainless steels in chemical environments intended to simulate in-reactor conditions have not positively identified any particularly severe chemomechanical degradation, such as that accompanying stress corrosion cracking or liquid metal embrittlement. We now report the observation of rapid, severe embrittlement of AISI 316 stainless steel by liquid tellurium-caesium mixtures in the temperature region 500–700 °C. Our results were obtained using a test technique that involves radially deforming small ring specimens contained inside an environmental chamber by compressive loading at a uniform displacement rate. This ring compression test technique has the advantages of facile control of the chemical environment and ease of adaptation for work with irradiated materials.

AISI 316 is an austenitic iron-base alloy containing by weight about 17% Cr, 13% Ni, 2.5% Mo, 2% Mn, 0.5% Si, 0.06% C and various trace elements. The liquid Te-Cs mixtures, which were applied to the alloy specimens as thin coherent coatings, contained dissolved oxygen at very low activity levels. The ring specimens used in our analyses were short (6.35 mm) lengths of 5.84 mm o.d., 0.381 mm wall AISI 316 seamless tubing with average grain diameters in the range 25–36 μm . Specimens, neutron-irradiated to a fluence of 8×10^{22} neutrons cm^{-2} ($E > 0.1$ MeV) at 488 °C in the Experimental Breeder Reactor-II at Idaho, were also tested.

Photomicrographs and scanning electron fractographs of unirradiated 20% cold-worked AISI 316 specimens stressed and broken in the presence of equimolar Te-Cs mixtures at 550–625 °C are shown in Figs 1–3. Figure 1 indicates the extreme brittleness of the fracture obtained at 625 °C, as illustrated by the sharp drop in load-bearing capability just past the yield load. The scanning electron micrographs of the 625 °C fracture surface shown in Figs 2b and 3a indicate the totally

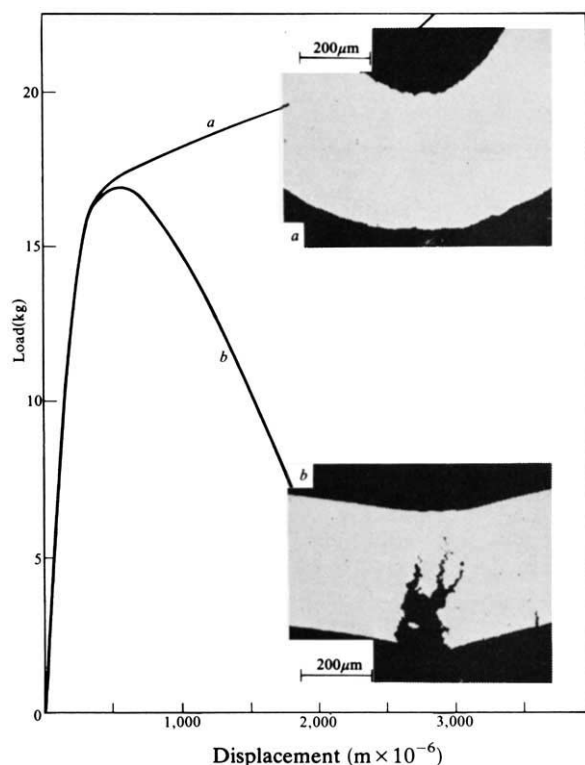


Fig. 1 Load-displacement curves and photomicrographs of unirradiated 20% CW AISI 316 ring specimens deformed at 625 °C in the presence of argon (a) and equimolar Te-Cs mixture (b); linear displacement rate $10.2 \mu\text{m s}^{-1}$ (peak strain rate $0.07\% \text{ s}^{-1}$).

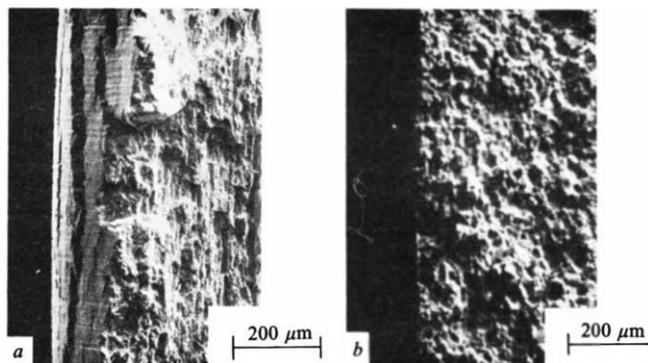


Fig. 2 Low magnification scanning electron fractographs of unirradiated 20% CW AISI 316 specimens stressed and broken in the presence of equimolar Te-Cs at 550 °C (a) and 625 °C (b); the direction of fracture is left to right for each specimen.

intergranular nature of the nonductile fracture. Similar results have been obtained on solution-treated AISI 316 and the fast neutron-irradiated 20% cold-worked materials. The embrittlement appeared to be even more marked in the irradiated material, occurring throughout the temperature range 450–625 °C with a peak load consistently lower than the load corresponding to initial yielding.

In equivalent test conditions, caesium alone did not embrittle cold-worked AISI 316, neither did simulated fission product mixtures of cadmium, indium, tin, antimony and iodine, with or without (excess) caesium. Liquid tellurium caused some reduction in yield stress at 625 °C, but no observable loss of ductility at either 475 °C or 625 °C. The marked embrittlement caused by liquid Te-Cs mixtures seems to be restricted to compositions lying between Cs_2Te and Te and, due to the absence of any concomitant Cs, Te-assisted oxidation of the alloy specimens, it is assumed to proceed independently of this corrosion process.

The observed 'hot' embrittlement of Type 316 stainless steel by liquid Cs-Te exhibits many features characteristic of 'classical' liquid metal embrittlement³. Only small quantities of the embrittling agent are required but it must be liquid and must be in intimate contact with the test specimen during loading. In one test, after initially coating the specimen with the Cs-Te mixture corresponding to the conditions for the test results given in Fig. 1, the coating was removed with successive washes of water and absolute ethanol. The load-deflection pattern for this specimen was found to be identical to the uncoated material, indicating that the embrittlement effect requires the presence of a surface layer of the liquid metal—a typical feature of liquid metal embrittlement. The stressed specimen in all cases exhibits normal elastic deformation up to some critical strain but, once this threshold has been exceeded, nonductile fracture rapidly follows; and, at a given strain rate, embrittlement occurs within a well-defined range of temperature whose lower limit usually coincides with the melting temperature of the embrittling agent.

The unique feature of this particular embrittlement process is the apparent chemical synergism between tellurium and caesium as embrittling agents—a synergism that is manifested by a marked dependence of embrittling agent activity on tellurium chemical potential. On the other hand, the activity of the embrittling agent did not appear to be markedly influenced by oxygen chemical potential, at least within the relatively narrow range of values expected inside a fast breeder reactor oxide fuel pin. The concept of 'inert carriers' in liquid metal embrittlement has been discussed elsewhere³; however, in the present case the role of caesium is thought to be more complex, being also possibly related to the manner in which wettability and tellurium activity change as a function of Cs/Te ratio.

We believe that the present results are the first observation of synergistic 'hot' embrittlement of AISI 316 stainless steel by liquid Te-Cs mixtures. This particular chemical synergy

parallels the synergy operating in Cs, Te-assisted oxidative corrosion of AISI 316 that sporadically appears inside fast reactor oxide fuel pins⁴, however, the mechanisms and consequences seem to be remarkably different.

Our recent results may have important implications for the design and operation of advanced fast reactor fuel pins. The composition and chemical states of fission products deposited on the inside surface of a fast reactor oxide fuel pin cladding are influenced by processes such as transport under thermal gradients, selective chemical reactions between fission products and between fission products and the fuel, and the temperature at a given location. Preliminary thermochemical analyses of the

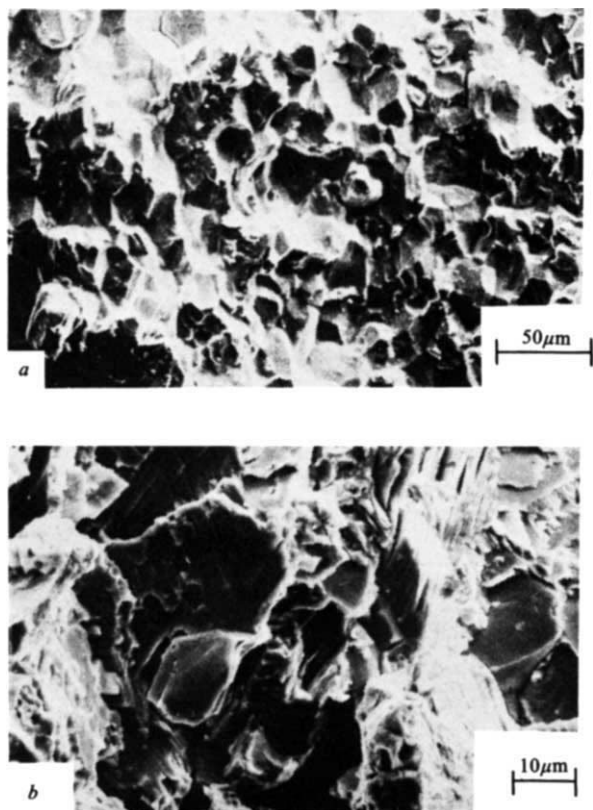


Fig. 3 High magnification scanning electron micrographs of intergranular (a) and transgranular (b) fracture surfaces from unirradiated 20% CW AISI 316 embrittled by equimolar Te-Cs; the corresponding test temperatures were 625 °C (a) and 550 °C (b).

conditions inside operating fast reactor fuel pins indicate that Cs/Te ratios <2 are possible at the fuel-cladding interface. However, for the specific mechanism discussed here to be operative in a particular fuel pin, the loading on the cladding would have to produce the specific threshold strain for liquid metal embrittlement at the same time that the chemical conditions are satisfied. Estimates of the associated probability will have to await the outcome of detailed chemical and mechanical modelling studies.

We thank T. A. Lauritzen and W. H. Reineking for experimental assistance, the US Department of Energy for financial support, and personnel at the Hanford Engineering Development Laboratory for supplying irradiated stainless steel specimens.

Received 20 July; accepted 30 October 1981.

1. Elliot, D. *Process Engng* 67-71 (July 1976).
2. Cahn, R. W. *Nature* 265, 11-12 (1977).
3. Old, C. F. *J. nucl. Mater.* 92, 2-25 (1980).
4. Adamson, M. G. *IAEA Rep. IWGFR/16*, 108-136, 170-189 (1979).

Carotenoid transformations in coastal marine waters

Daniel J. Repeta & Robert B. Gagosian

Department of Chemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

Rapidly sinking large particles contribute a major portion of the mass flux from marine surface waters to the sea floor. The recognition that organic matter associated with these particles undergoes major alterations in its chemical composition during vertical transit has raised new questions on the role of rapid transformation processes in carbon and nitrogen cycling, benthic nutrition, and post-depositional degradation¹. Analysis of specific organic compounds has established some of the compositional changes of these transformations²⁻⁴. However, most of the changes observed represent the sum of biological production and degradation. For example, bacterial and chemical degradation of fatty acids in the water column may be completely masked by an even greater production of these compounds by zooplankton. Hence it is often difficult to distinguish between competing degradative processes, and to assess the relative significance of each. Carotenoid pigments have characteristics which make them particularly suitable for use as organic tracers to distinguish between short-term degradation processes. Carotenoids are widely distributed in both photosynthetic and nonphotosynthetic organisms; many of them are source specific; they contain diverse functional groups; and they represent one of the more labile biogenic compound classes in the sea. We report here for the first time the isolation and identification of carotenoid transformation products from sediment trap experiments conducted in coastal marine waters, and discuss the source and significance of observed transformation products with respect to a proposed degradation pathway.

Sediment traps were deployed in Buzzards Bay, Massachusetts (41°32' N, 70°42' W) for two 30-day periods between 20 May and 23 July 1980. Traps were set 4 m below the surface in a 10-m water column. Collected material was filtered and sonic-extracted with methanol (twice, 20 min each) and dichloromethane (once, 20 min). Extracts were concentrated

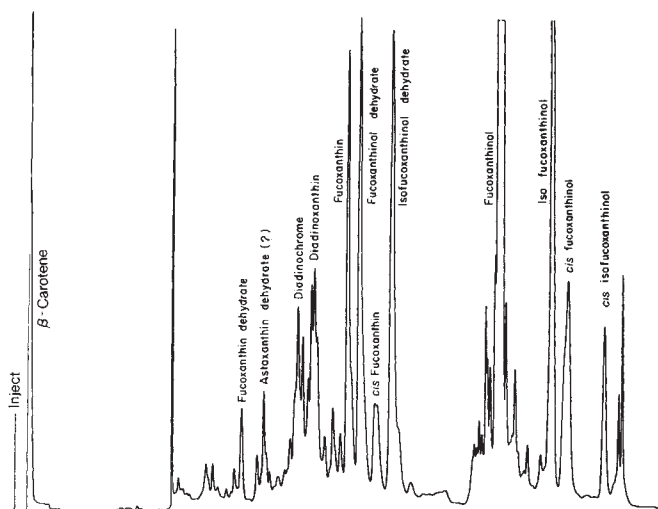


Fig. 1 HPLC trace of Buzzards Bay sediment trap sample no. 3. Conditions are: 30×3.9 mm 5 μm amino column, eluted with a linear gradient of 0-13% THF/methanol (80/20, v/v) in hexane at 2 ml min⁻¹ for 45 min. After 55 min the eluant was stepped to 30% THF/methanol for an additional 10 min. Detection was at 436 nm.

Table 1 Fucoxanthin metabolites in Buzzards Bay sediment trap 3, 21 June–23 July 1980

	% Total*	% Phytoplankton
Fucoxanthin (Ia)	5	100 ⁷
Fucoxanthinol (Ib)	69	Trace ⁸
Isfucoxanthinol (IIb)	12	ND
Fucoxanthin dehydrate (IIIa)	1	ND
Fucoxanthinol dehydrate (IIIb)	6	ND
Isfucoxanthinol dehydrate (IV)	7	ND
	100%	

ND, not detected.

* Total fucoxanthin-derived pigments, relative abundances not corrected for small differences in extinction coefficients (see ref. 9).

under reduced pressure then separated into compound classes by gel permeation chromatography using 100 Å μ Styragel (Waters Associates) with dichloromethane as the mobile phase. This procedure removes low-molecular weight lipids and chlorophyll pigments which interfere with subsequent steps in the analysis. The carotenoid fraction was collected, re-concentrated, then separated into individual components by high pressure liquid chromatography (HPLC) using a Spherisorb 5 μ m amino column (300 $\times$ 3.9 mm) eluted with a linear gradient of hexane and 0–13% THF/methanol (80/20, v/v) at 2 ml min⁻¹ (Fig. 1). Carotenoids were detected spectroscopically at 436 nm, collected, and further analysed by visible spectroscopy and mass spectrometry. Mass spectra were collected on a Finnigan 3200 quadrupole spectrometer interfaced with an INCOS (Finnigan) 2300 data system. Conditions are described in Fig. 2 legend. All operations were performed at or below 20 °C and in low light conditions. A more detailed description of the analytical method will be given elsewhere.

The lability of carotenoids has posed severe problems in their analysis⁵. When discussing geochemical transformations in particular, care must be taken to ensure transformation products are not artefacts of the analytical procedure. Analysis of blanks spiked with carotenoid mixtures (astaxanthin, β -carotene, diadinoxanthin, fucoxanthin and lutein), and of cultured phytoplankton (*Peridinium trichoidium*) of known carotenoid composition⁶ showed that no degradation other than *cis* $\rightleftharpoons$ *trans* isomerization occurs during our analytical work-up.

Fucoxanthin (Ia) is the major carotenoid in diatoms and a few dinoflagellates, often constituting 70–90% of the total carotenoids⁷. In our sediment trap samples, fucoxanthin comprised only 4% of the total carotenoid and 5% of the total fuco-pigments (Table 1, Fig. 1). The remaining 95% of the total fuco-pigments consisted of the fucoxanthin transformation products fucoxanthinol (Ib), isofucoxanthinol (IIb), fucoxanthinol dehydrate (IIIb), isofucoxanthinol dehydrate (IV) and fucoxanthin dehydrate (IIIa). Together these compounds constitute 82% of the total carotenoids with diadinoxanthin (V) and diadinochrome (VI) contributing another 3% and 1%, respectively. Fucoxanthinol (Ib) has been reported as a minor component of some algae, and is thought to be an intermediate in fucoxanthin biosynthesis^{8,9}. Isofucoxanthinol (IIb) is known to arise from base-catalysed epoxide opening and is commonly cited as an analytical artefact¹⁰. For reasons outlined above, we believe that the opened epoxides (iso-compounds) in these samples are not artefacts of the analytical method, but are due to chemical transformations in the water column. We know of no reports of naturally occurring dehydrates (IIIa, b) of either fucoxanthin or fucoxanthinol.

Fucoxanthinol (Ib) (Fig. 2) is the most abundant pigment in the samples (Table 1). Acylation with acetic anhydride in pyridine yields fucoxanthin acetate which co-elutes with standard fucoxanthin acetate synthesized from fucoxanthin (see Fig. 2 legend, which gives mass spectrometry and IUPAC data for all compounds discussed in this paper). Thermal isomerization of fucoxanthin acetate produces an equilibrium mixture of *cis* isomers identical in both chromatographic retention time (HPLC) and relative abundance to those derived from the

standard. The ester hydrolysis which converts fucoxanthin to fucoxanthinol can occur by two pathways, a base-catalysed chemically mediated ester hydrolysis, or biochemical metabolism. Base-catalysed ester hydrolysis seems unlikely as epoxide opening should precede ester hydrolysis¹⁰. In our samples fucoxanthinol, which still contains the epoxide group, predominates over isofucoxanthinol, its open epoxide counterpart. Biochemical metabolism by zooplankton and other higher organisms seems a more likely source. Fucoxanthinol has been found as a major fuco-pigment in the guts of sea urchins¹¹, thus it is not unreasonable that other higher organisms have the ability to hydrolyse the ester without rearranging the epoxide. As particles collected in sediment traps are thought to represent material which has been extensively reworked by heterotrophic

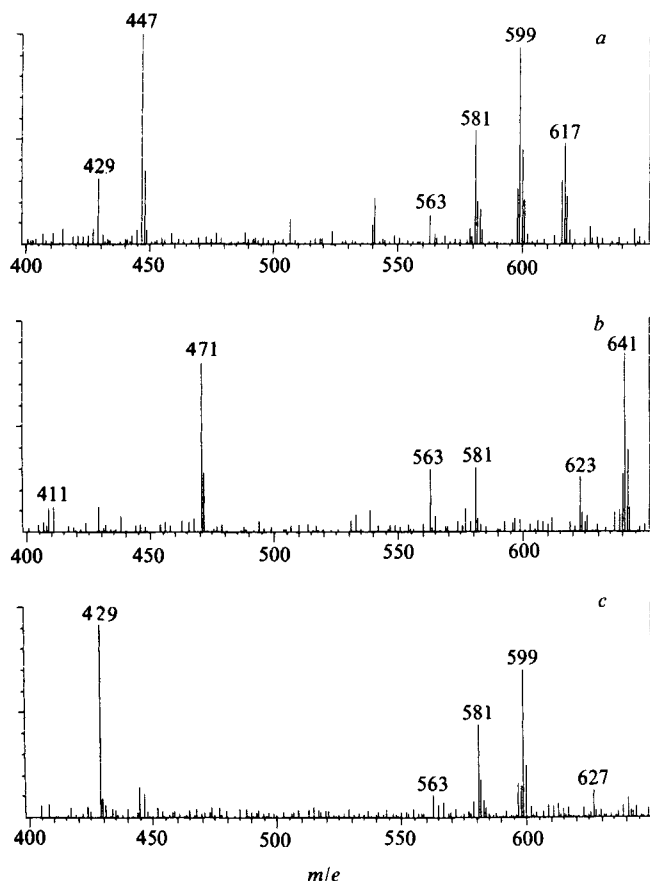


Fig. 2 Chemical ionization mass spectra of *a*, fucoxanthinol (Ib) ($M+1=617$, $M+1-18(-H_2O)=599$, $M+1-18-18=581$, $M+1-18-18-18=563$, $M+1-170$ (cleavage to the 7,8 ketone)=447, $M+1-170-18=429$); *b*, fucoxanthin dehydrate (IIIa) ($M+1=641$, $M+1-18=623$, $M+1-60(-acetate)=581$, $M+1-18-60=563$, $M+1-170=471$); and *c*, fucoxanthinol dehydrate (IIIb) ($M+29=627$, $M+1=599$, $M+1-18=581$, $M+1-18-18=563$, $M+1-170=429$); CH_4 reactant gas 900 μ m, ionization voltage 130 eV, ionization current 500 μ A. Where: fucoxanthinol (Ib) (5,6-epoxy-3,3',5'-trihydroxy-6',7'-didehydro-5,6,7,8,5',6'-hexahydro- β,β caroten-8-one) ($M+1=617$, $M+1-18=599$, $M+1-18-18=581$, $M+1-18-18-18=563$, $M+1-170-18=429$). Fucoxanthin acetate (5,6-epoxy-3,3',5'-trihydroxy-6',7'-didehydro-5,6,7,8,5',6'-hexahydro- β,β caroten-8-one, 3,3'-diacetate) ($M+1=701$, $M+1-18=683$, $M+1-60=641$, $M+1-18-60=623$). Fucoxanthin dehydrate (IIIa) (5,6-epoxy-3,3'-dihydroxy-4',6',7'-tridehydro-5,6,7,8,6'-pentahydro- β,β caroten-8-one 3'-acetate) ($M+1=641$, $M+1-18=623$, $M+1-60=581$, $M+1-18-60=563$, $M+1-170=471$). Fucoxanthinol dehydrate (IIIb) (5,6-epoxy-3,3'-dihydroxy-4',5',7'-tridehydro-5,6,7,8,6'-pentahydro- β,β caroten-8-one) ($M+1=599$, $M+1-18=581$, $M+1-170=429$). Dehydro-fucoxanthin diacetate (5,6-epoxy-3,3'-dihydroxy-4',6',7'-tridehydro-5,6,7,8,6'-pentahydro- β,β caroten-8-one 3,3'-diacetate) ($M+1=683$, $M+1-60=623$, $M+1-212=471$). Double bond in dehydrates tentatively assigned to the 4',5' position. (See ref. 10.)

organisms¹, we propose that fucoxanthinol represents a metabolite of heterotrophic activity. Other studies conducted in our laboratory, concerned with the direct collection of zooplankton faecal pellets from field cultures, verify this hypothesis¹².

The mass spectra of fucoxanthin dehydrate (IIIa) and fucoxanthinol dehydrate (IIIb) (Fig. 2) are both consistent with loss of the tertiary alcohol at the 5' position (Fig. 3). The mass fragment $M + 1 - 170$ arises from a cleavage of the 7,8 bond α to the 8-ketone. The resulting mass fragments (471, 429 AMU) correspond to a loss of a hydroxyl group at either the secondary 3' or tertiary 5' position. Acylation of the isolated fucoxanthinol dehydrate (IIIb) yields a diacetate. Acetic anhydride in pyridine acylates only primary and secondary alcohols in the conditions used in our synthesis. The mass increase of 84 AMU for the acylated dehydrate over the underivatized parent is consistent with the esterification of the two hydroxyl groups at the 3 and 3' position. Therefore, dehydration proceeds with the loss of the 5' alcohol. Sufficient amounts of fucoxanthin dehydrate (IIIa) could not be isolated to derivatize, hence our conclusion of 5' hydroxyl loss for this compound rests on mass spectral evidence alone.

Determining the mechanism of dehydration is difficult. There is little evidence in favour of either a biochemically or chemically mediated reaction. Chemically, an acid-catalysed dehydration reaction seems most likely. However, we find substantial amounts of the unrearranged 5,6 epoxide, diadinoxanthin (V) in our samples. In acidic conditions the 5,6 epoxide rapidly rearranges to the 5,8 furanoxide, diadinochrome (VI)¹³. If the acidic conditions necessary to dehydrate fucoxanthin existed in our samples, all the diadinoxanthin should have rearranged to the 5,8 furanoxide. Although some diadinochrome was observed, we believe that dehydration is probably not a result of chemical acid catalysis, but a more site-specific metabolic reaction.

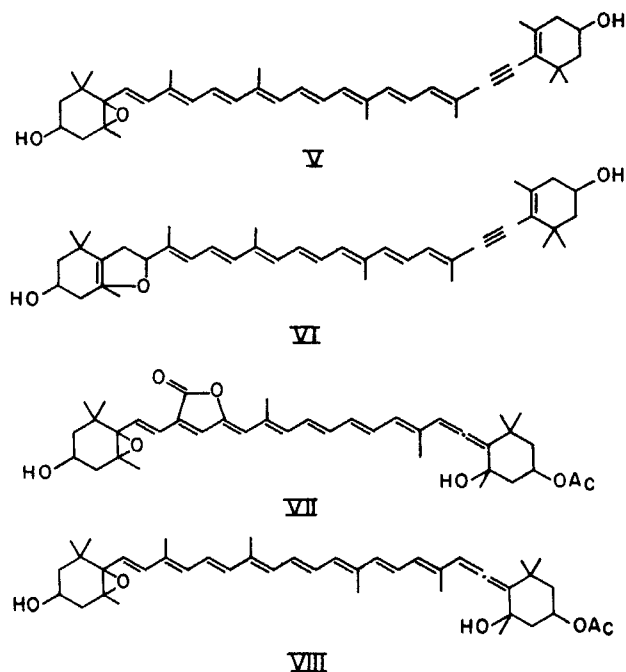


Fig. 4 Molecular structure of diadinoxanthin (V), diadinochrome (VI), peridinin (VII) and dinoxanthin (VIII).

The dehydration of sterols^{15,16}, amino acids¹⁷, and phytol¹⁸ occurs in a variety of marine environments and both chemical and biochemical mechanisms have been proposed. The dehydration of stanols to sterenes is structurally analogous to the dehydration of the secondary 3 or 3' hydroxyl of carotenoid alcohols. The dehydration products of fucoxanthin (Fig. 3, IIIa, b) arise from loss of the tertiary 5' hydroxyl group, with retention of the hydroxyl groups at both the 3 and 3' positions. This suggests either that different processes exist for the dehydration of stanols and carotenoid alcohols, or there is a general process which displays an order of biochemical reactivity (for example, tertiary alcohols dehydrate faster than secondary alcohols). In the latter case we would expect to observe a sequential dehydration through the 3,5' (or 3',5') didehydro and 3,3',5' tridehydro intermediates. The carotenoid mixtures in our samples contain ~100 components, of which only the 20 major components have been identified by our methods. Our analysis does not exclude the existence of more highly dehydrated metabolites in our samples.

In conclusion, we have observed ester hydrolysis and dehydration of fucoxanthin associated with particles collected in sediment traps. We propose a biochemically mediated transformation pathway (Fig. 4) which predicts that other carotenoids structurally similar to fucoxanthin (such as peridinin (VII) and dinoxanthin (VIII) from dinoflagellates) should undergo analogous transformations. We are conducting experiments with both long- and short-term sediment trap deployments, and direct collection of zooplankton faecal pellets to substantiate the proposed mechanisms.

This research was supported by the NSF, Division of Ocean Sciences, grant nos OCE 79-25352 and OCE 81-18436 and by contract no. N00014-74-CO-262 NR083-004 from the Office of Naval Research. D.J.R. thanks the Woods Hole Oceanographic Institution for financial assistance. We thank Drs Nelson M. Frew for mass spectrometric analyses, John W. Farrington, Cindy Lee and John K. Volkman for valuable comments, and Tim Cowles for cultured phytoplankton. Woods Hole Oceanographic Institution contribution no. 4957.

Received 19 August; accepted 3 November 1981.

1. Honjo, S. *J. mar. Res.* **38**, 53-97 (1980).
2. Wakeham, S. G. *et al. Nature* **286**, 798-800 (1980).
3. Lee, C. & Cronin, C. C. *J. mar. Res.* (in the press).

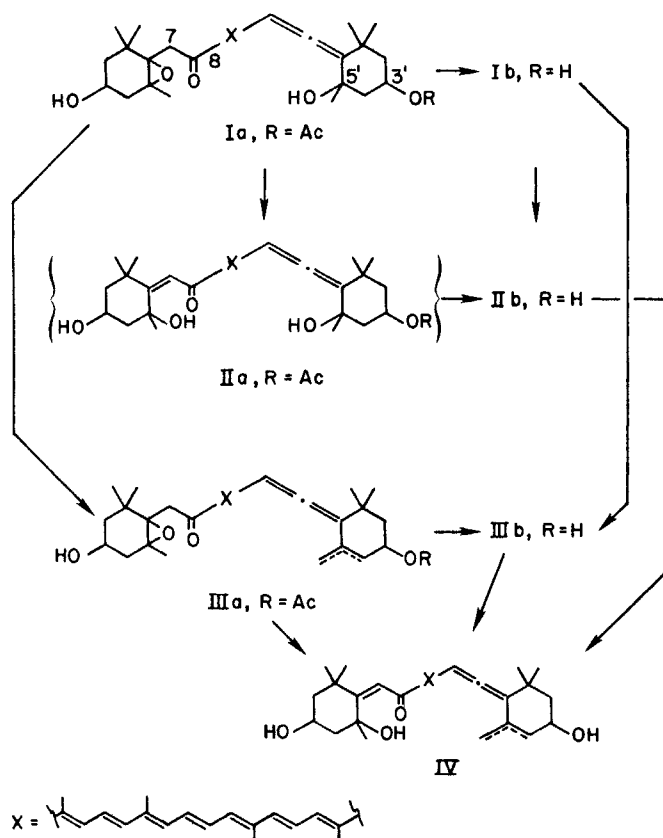


Fig. 3 Proposed pathway showing ester hydrolysis, dehydration and epoxide opening for biochemical and chemical degradation of fucoxanthin. Compound in parentheses was not isolated in this study.

4. Gagosian, R. B., Smith, S. O. & Nigrelli, G. E. *Geochim. cosmochim. Acta* **45**, 2161–2168 (1981).
5. Davies, B. H. in *Chemistry and Biochemistry of Plant Pigments* Vol. 2 (ed. Goodwin, T. W.) 38–165 (Academic, New York, 1971).
6. Johansen, J. E., Svec, W. A. & Liaaen-Jensen, S. *Phytochemistry* **13**, 2261–2271 (1974).
7. Goodwin, T. W. (ed.) in *Aspects of Terpenoid Chemistry and Biochemistry* 315–356 (Academic, New York, 1971).
8. Liaaen-Jensen, S. in *Marine Natural Products* Vol. 2 (ed. Scheuer, P. J.) 2–73 (Academic, New York, 1978).
9. Liaaen-Jensen, S. in *Marine Natural Products Chemistry* (eds Faulkner, D. J. & Fenical, W. H.) 239–260 (Plenum, New York, 1977).
10. Bonnett, R. *et al. J. chem. Soc. C*, 429–454 (1969).
11. Galasko, G. *et al. J. chem. Soc. C*, 1264–1265 (1969).
12. Repeta, D. J. & Gagosian, R. B. in *Advances in Organic Geochemistry* (ed. Bjoroy, M.) (Heyden, London, in the press).
13. Buchecker, R. & Liaaen-Jensen, S. *Phytochemistry* **16**, 729–733 (1977).
14. Gagosian, R. B., Smith, S. O., Lee, C., Farrington, J. W. & Frew, N. M. in *Advances in Organic Geochemistry 1979* (eds Douglas, A. G. & Maxwell, J. R.) 407–419 (Pergamon, Oxford, 1980).
15. Dastillung, M. & Albrecht, P. *Nature* **269**, 678–679 (1977).
16. Bada, J. L. & Hoopes, E. A. *Nature* **282**, 822–823 (1979).
17. Simoneit, B. R. *Init. Rep. DSDP* **21**, 909–923 (1973).

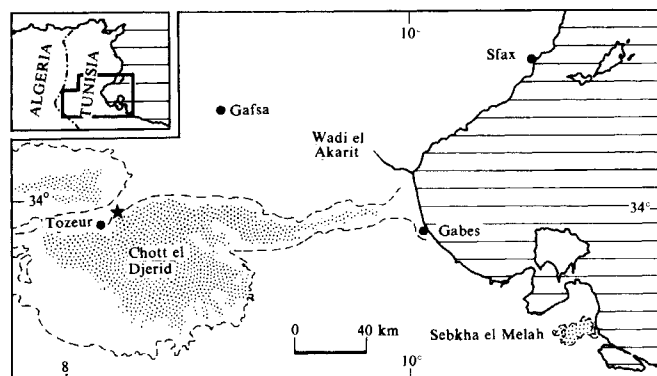


Fig. 1 Location map. Dashed line (after ref. 3) shows former extent of Chott el Djerid. Stipple indicates present-day salt-covered depressions. Star shows location of sites sampled for ^{14}C dating.

Marine deposits 35,000–25,000 years old in the Chott el Djerid, southern Tunisia

G. W. Richards & C. Vita-Finzi

Department of Geography, University College, Gower Street, London WC1E 6BT, UK

The Chott el Djerid is a large, salt-covered depression in southern Tunisia (Fig. 1). The late nineteenth century idea that the basin had been invaded by the sea^{1,2} has been superseded by the view that any Quaternary flooding of the Chott must have resulted from climatic changes, as the fossils found in its sediments are not exclusively marine and there is no indication of the fluctuations in sea level or of the crustal movements that would be required to fill and then empty the basin^{2,3}. We report here evidence for a marine connection 35,000–25,000 years ago and show that it implies subsequent uplift by some 80 metres in an area long regarded as either stable² or prone to subsidence throughout the Quaternary⁴.

Two sets of terraces border the Chott. The earlier terrace dates from Villafranchian (Lower Pleistocene) times. The younger terrace contains numerous shells of the bivalve *Cerastoderma* (= *Cardium*) *glaucum* (Bruguière) and a few of the gastropods *Melania tuberculata* (Müller) and *Hydrobia* sp. In the Tozeur area *Corbicula fluminalis* (Müller) is also present. The microfauna includes 14 species of marine foraminifera, predominantly *Ammonia beccarii* (d'Orb.), and three species of ostracoda (*Cyprideis* sp.)⁵.

Radiocarbon ages ranging from $22,000 \pm 800$ to $32,400^{+4,100}_{-2,800}$ yr BP on shell samples obtained from the younger terrace by Page⁶ supported previous estimates² of its age and confirmed that it formed at a time when sea level was lower than today. Flooding was thus ascribed to climatic oscillations whether in the guise of reduced evaporation and thence a rise in the water table³ or of an increase in rainfall effectiveness⁶. The fauna appeared to be consistent with lacustrine conditions. *C. glaucum* can tolerate salinities of 3–60‰ (ref. 7). *M. tuberculata*, though sometimes washed onto modern beaches (as at Gafes), is essentially a freshwater species. Both species occur together with *Hydrobia* sp., *Corbicula africana* and other freshwater or

brackish water organisms as well as Foraminifera in Quaternary lake deposits in Wadi Shati, south Libya^{8,9}.

In March 1981 we found numerous rolled fragments of the oyster *Ostrea stentina* (Payr.) in one of the two type sections³ in the younger terrace at Helba, 1.5 km north-east of Tozeur (Fig. 1) and 40 m above sea level. The deposits, which also contained abundant *C. glaucum* and a few specimens of *M. tuberculata*, consisted of a pebbly sand 30-cm thick resting unconformably on a southward-dipping clastic series ascribed to the Plio-Pleistocene³. What appears to be a regressive shoreline is exposed 5 m lower and some 200 m to the south-east of the first section. Here *O. stentina* is present within two gravel beds separated by 30 cm of massive sand containing a few *C. glaucum*. The upper bed is 5 cm thick and composed of large pebbles, the lower is a fine gravel at least 20 cm thick (Fig. 2).

Ostrea stentina is now found living in the intertidal zone at Gafes and elsewhere on the Tunisian coast. Unlike *C. glaucum* it has not been recorded from hypersaline or brackish lagoons¹⁰. Redeposition from older rocks is unlikely because the only oysters hitherto reported from the local Tertiary rocks are *O. multicostata* Desh., *O. punica* Thomas and *O. bellovacina* Lmk¹¹. Transport by man cannot be ruled out, but it is improbable in view of the absence of complete shells and of artefacts or other evidence of human occupation at the two sections, and the lack of oyster shells from local prehistoric sites¹².

Radiocarbon ages were determined on oysters from the two exposures (Fig. 2) and also on a sample of *C. glaucum* from the upper exposure. No sample showed evidence of recrystallization when examined in thin section, and X-ray diffraction confirmed that the *C. glaucum* shells had retained their aragonitic composition. The stable isotope measurements (Table 1) and the good agreement with the ^{14}C ages previously reported⁶ for the terrace are consistent with this view¹³. If one accepts that ~30,000-yr ago sea level lay 40 m below its present position¹⁴, the beach deposits represent total emergence by 80 m and 75 m respectively, and hence uplift rates of 2.3–3.3 mm yr⁻¹. A sea level of -8 m as indicated by the sedimentary record beneath the Sebkhah el Melah¹⁵ (Fig. 1) would yield uplift rates of 1.4–1.7 mm yr⁻¹.

The results conflict with the evidence presented by geomorphologists for local stability² and by geologists for local sub-

Table 1 Sample data from the Chott el Djerid

Sample no.	Lab. no.	Lat. and long.	Age* (yr BP)	$\delta^{13}\text{C}$ ‰	$\delta^{18}\text{O}$ ‰	Species
T81/2o	BETA-2653	33°56' N, 8°9' E	35,140 ± 1,190	1.14	-2.67	<i>Ostrea stentina</i>
T81/2c	BETA-2615	33°56' N, 8°9' E	28,950 ± 280	-0.06	-1.17	<i>Cerastoderma glaucum</i>
T81/3	BETA-2654	33°56' N, 8°9' E	25,360 ± 270	1.27	-2.51	<i>Ostrea stentina</i>

* Normalized to $\delta^{13}\text{C} = 0\text{‰}$ PDB.

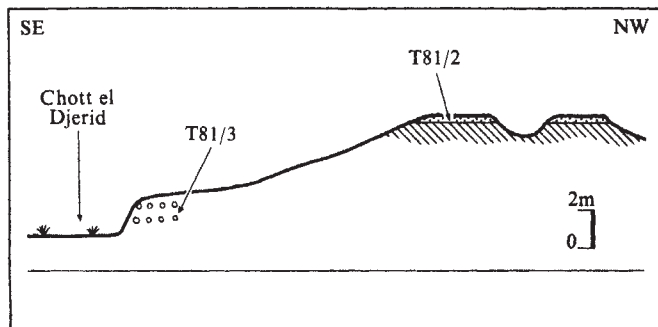


Fig. 2 Section at sample locality. Stipple represents the older of the two dated terraces; open circles denote gravel bands in younger terrace; inclined lines denote Plio-Pleistocene beds. The two sample points are listed in Table 1.

sidence⁴, but accords with reports of uplift by 9 m since 9,500 yr BP (equivalent to an average of 3.9 mm yr⁻¹) in the Wadi el Akarit⁶ (Fig. 1). It remains to be seen whether other parts of the Chott el Djerid yield a similar record and how it relates to the Pleistocene folding at Gafsa¹⁶ and the slight

Holocene uplift detected on the Tunisian coast south-east of Jerba¹⁷.

We thank John Billeter, Solene Morris, Pat Nuttall, Reza Samadian and Brian Walker for their help. G.W.R. holds a NERC studentship.

Received 9 June; accepted 20 November 1981.

1. Lesseps, F. de *Rev. Scient.* **21**, 495–496 (1883).
2. Coque, R. *La Tunisie Préaharienne* (Colin, Paris, 1962).
3. Coque, R. & Jauzein, A. in *Guidebook to the Geology and History of Tunisia*, 227–257 (Petroleum Exploration Society of Libya, Tripoli, 1967).
4. Burollet, P. F. *Annls Mines Géol. Tunis* **18**, 1–345 (1956).
5. Milokhoff, E. *C.r. XIXe Congr. géol. int., Alger* **15**, 337–339 (1954).
6. Page, W. D. thesis, Univ. Colorado (1972).
7. Gaillard, J. M. & Testud, A.-M. in *The Geology of Libya* Vol. 3 (eds Salem, M. J. & Buswail, M. T.) 809–813 (Academic, London, 1980).
8. Gaven, C., Hillaire-Marcel, C. & Petit-Maire, N. *Nature* **290**, 131–133 (1981).
9. Petit-Maire, N., Casta, L., Delibrias, G., Gaven, C. & Testud, A.-M. in *The Geology of Libya* Vol. 3 (eds Salem, M. J. & Buswail, M. T.) 797–807 (Academic, London, 1980).
10. Zaouali, J. *Bull. off. nam. Pêch. Tunisie* **3**, 107–140 (1979).
11. Domergue, Ch., Dumon, E., De Lapparent, A. F. & Lossel, P. *C.r. XIXe Congr. géol. int., Alger, Monogr. rég. Tunisie* **7** (1952).
12. Gobert, E. G. *Quaternaria* **6**, 271–307 (1962).
13. Millman, J. D. *Marine Carbonates* (Springer, New York, 1974).
14. Chappell, J. & Veeh, H. H. *Nature* **276**, 602–603 (1978).
15. Fontes, J.-C. & Perthuisot, J.-P. *Nature* **244**, 74–75 (1973); *Rev. géol. Phys. géol. Dyn.* **13**, 299–314 (1971).
16. Castany, G. *Geol. Rdsh.* **43**, 196–203 (1955).
17. Dalongeville, R., Paskoff, R., Sanlaville, P., Thommeret, J. & Thommeret, Y. *C.r. hebdom. Séanc. Acad. Sci., Paris D290*, 303–306 (1980).

Myoblasts from *Drosophila* wing disks can contribute to developing muscles throughout the fly

Peter A. Lawrence & Danny L. Brower*

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

In insect embryos groups of epidermal cells are already determined¹ to form particular compartments of the adult^{2–4}. In *Drosophila* this determination is maintained through cell divisions, is stable to experimental interference^{1,5,6} and is thought to depend on the activity of a particular class of homoeotic genes^{4,7,8}. In contrast to our understanding of the ectoderm, we know little about determination of the mesoderm, which gives rise to the musculature. A recent study of cell lineage in the thorax of adult *Drosophila* has shown that the muscles of each segment are divided precisely into sets (P.A.L., in preparation). The dorsal and ventral muscle sets of the mesothorax probably derive from ad epithelial myoblasts^{9–11} in the wing and leg imaginal disk, respectively (P.A.L., in preparation). We describe here a new method of marking cells to trace myoblasts from the wing imaginal disk. We show that, when implanted, these mesodermal cells do not develop as they would *in situ* but instead contribute indiscriminately to developing host muscles. This result suggests that, whereas the ectodermal cells of the disk are committed to form a particular region of the adult, the mesodermal cells are not.

Portions of the wing imaginal disk from a wild-type larva were implanted into mature larval hosts. The donor cells carried the wild-type allele (*sdh*⁺) at a locus which affects the enzyme succinate dehydrogenase; the host larvae were mutant (*sdh*²/*sdh*⁷) at this locus¹². After metamorphosis the flies, as well as unimplanted controls of both genotypes, were histochemically stained for enzyme activity (see Fig. 1 legend). In the *sdh*⁺ flies all muscles stained an intense dark blue, the stain being localized in the mitochondria. In the *sdh*²/*sdh*⁷ flies most muscles varied between pure white and pale blue in colour; 25 abdomens were examined, and dark blue muscles were

seen only in the genital region. For this reason the genital muscles were not studied in the following experiments. Of 25 mutant thoraces sectioned, all muscles were white to pale blue.

A portion of the wild-type wing disk (02 piece¹³; see Fig. 1) containing muscle precursors, the ad epithelial cells, was implanted into 29 adults. In every case where the implant was retrieved it consisted of mesonotum and wing hinge structures, as expected¹³, and was usually found in the posterior region of the abdomen. Only occasionally were implants associated with substantial numbers of adventitious muscle fibres and these could not be recognized as being of any particular type. Near the implant, some of the hosts showed abdominal muscles that were frequently (28/29 cases) stained dark blue (Fig. 2). These host muscles appeared normal in every other respect. The intensity of staining varied from fibre to fibre, densely stained fibres often being intermingled with unstained ones (Fig. 3). Further from the implant only occasional muscle fibres were stained, but these were prominent against the pale unstained background of the other fibres. The pleural membrane of the abdomen is covered with a delicate sheet of laterally oriented fibres, and these were also stained blue near the implant. In addition, the thoraces of many of the experimental flies (10/26) contained stained muscle fibres. For example, dark patches of stain were observed in both the direct and indirect flight muscles but were never found in the unimplanted controls (Fig. 4a, b).

There were three successful implantations of portions of the wing disk which probably do not contain ad epithelial cells¹⁴ (58 pieces, Fig. 1). These implants formed the expected type of cuticle (wing blade and thoracic pleura¹³) and were not associated with any muscle fibres. All the host muscles of the abdomen and thorax were normal in structure and position and none stained dark blue.

Clonal analysis has shown that *sdh* alleles are cell-autonomous markers¹² which can be used in the muscles (P.A.L., in preparation). It is therefore highly unlikely that the succinate dehydrogenase activity was rescued other than by myoblast fusion. This is supported by the absence of staining when pieces of disk (58) lacking ad epithelial cells are used. The sporadic staining of fibres far from the site of the implant also suggests that the myoblasts migrate and fuse with host muscles. The fact that pale, intermediate or dark fibres are closely intermingled suggests that none, a few or many donor cells may contribute to each fibre. It therefore seems that *sdh*⁺ myoblasts, released from the wing imaginal disk, fuse indiscriminately with host muscles.

* Present address: Developmental Biology Center, University of California, Irvine, California 92717, USA.

The adult muscles vary in many respects including sarcomere length, nuclear shape, mitochondrial size and fibre dimensions. However, when the host muscles of the abdomen and thorax stained blue, and were therefore at least partly composed of donor myoblasts, their structure appeared normal in every detail. There is evidence that the donor myoblasts *in situ* form only the dorsal muscle set of the mesothorax, which includes all the indirect flight muscles and a specific set of the direct flight muscles (P.A.L., in preparation). After implantation, they exhibit no such restriction and can apparently contribute to any host muscles with which they happen to fuse. Thus when removed from their normal environment, the ad epithelial cells are not determined with respect to the type and location of the muscles they will form. By contrast, the fate of the epidermal

cells is determined; when transplanted into mature larvae they differentiate according to their prospective fate^{1,13}.

In contrast to earlier studies^{15,16} it has been shown that the ad epithelial cells do not arise directly from the disk epithelium (refs 17, 18 and P.A.L., in preparation) and fate-mapping studies of the adult flight muscles place their precursors in the mesodermal region of the blastoderm, at some distance from the presumptive epidermis¹⁹. The ad epithelial cells can therefore be designated as mesoderm. Experiments on insect embryos have shown that when pieces of ectoderm are removed or damaged,

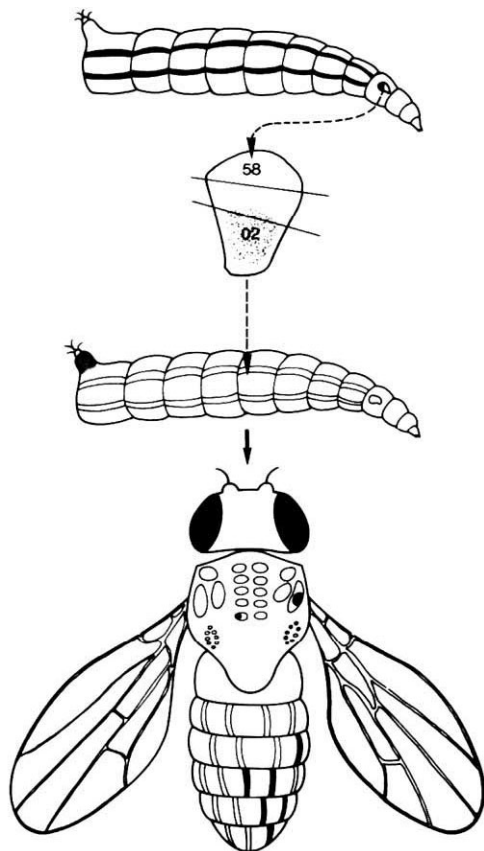


Fig. 1 Summary of the experimental procedure. A wing imaginal disk was dissected from a wild-type larva, and either the 58 or 02 piece¹³ was sliced with tungsten needles and transplanted into the abdomen of a mutant host larva. After metamorphosis the host fly was stained for enzyme activity; dark fibres were found sporadically in the muscles, especially near the site of the implant. The shaded area in the disk shows the distribution of ad epithelial cells, based on the binding patterns of fluorescent peanut lectin and a monoclonal antibody¹⁴. For marking the cells we used different alleles at the *sdh* locus¹². Hosts were *b sdh²/b sdh⁷* and therefore contain a thermolabile form of the enzyme succinate dehydrogenase¹². After careful heating the succinate dehydrogenase activity disappears in *sdh²/sdh⁷* muscles but survives in *sdh⁺* muscles. Donors were from a wild-type (Barton) stock. The *black* (*b*) mutation (see ref. 24) allowed the identification of the *sdh²/sdh⁷* host larvae, which could be distinguished from their siblings by the dingy colour of the posterior spiracles. Enzyme activity, which is located in mitochondria, was specifically stained using the reduction of nitrobluetetrazolium¹². Mature *sdh⁺* disk fragments were implanted into the abdomen of fully grown *sdh²/sdh⁷* larvae by standard methods²⁵. After metamorphosis the abdomen was dissected with the implant, heated to 52 °C in Ringer's solution for 15 min then stained for 2 h. The thoraces were heated to 53 °C in Ringer's solution for 15 min, stained for 2–3 days, embedded in paraffin wax and sectioned at 25 µm.

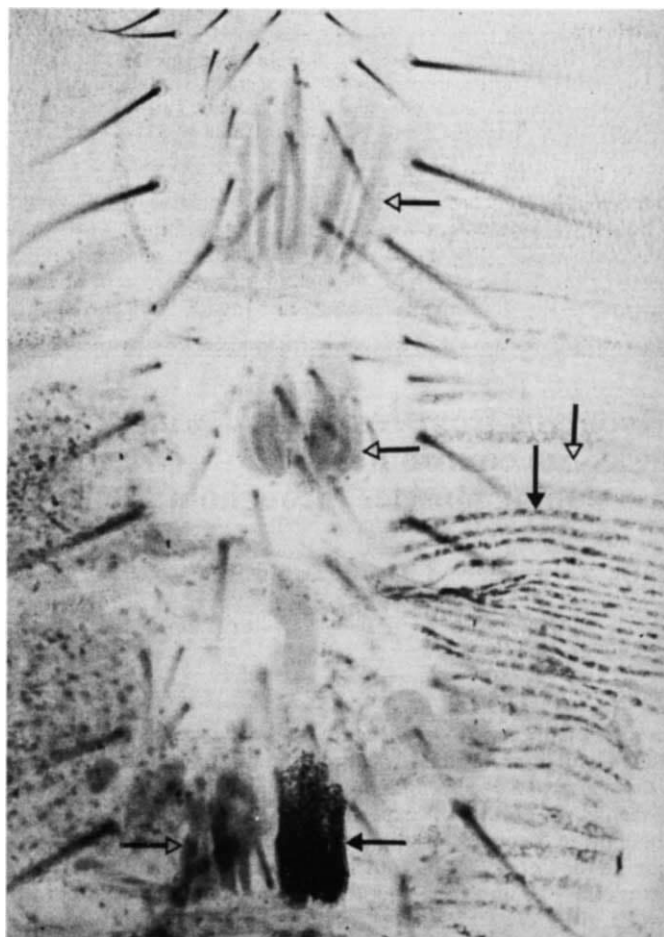


Fig. 2 The dissected ventral abdomen of a host after implantation of a 02 piece of wing disk. Note that in the vicinity of the implant, numerous muscle fibres have stained dark blue (solid arrows). Open arrows indicate unstained muscles. For description of the adult muscles of *Drosophila* see Miller²⁶. ×230.

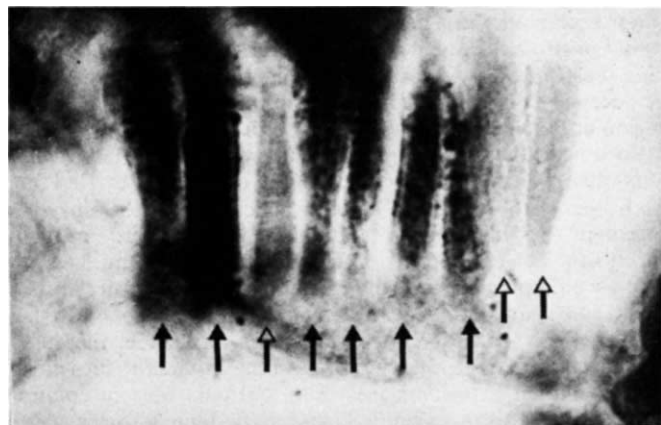


Fig. 3 Detail of abdominal muscles from another implant prepared as in Fig. 2 legend. Note that stained fibres (solid arrows) are intermingled with unstained ones (open arrows). ×1,000.

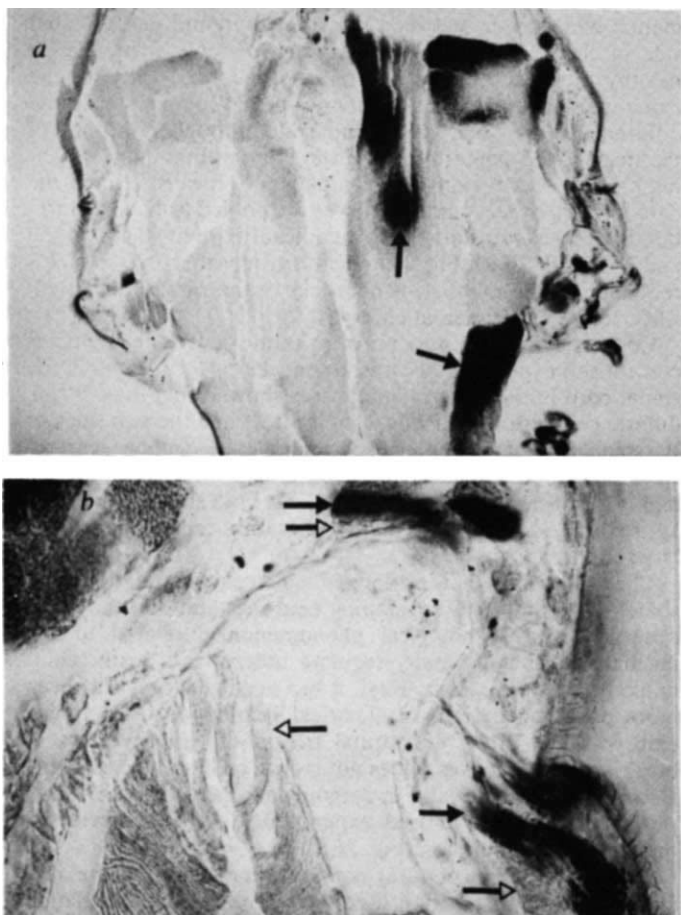


Fig. 4 *a*, Section of the thoracic muscles from an individual containing a *sdh*⁺ implant in the abdomen. The dark regions (solid arrows) in the indirect flight muscles indicate the participation of donor myoblasts. $\times 115$. *b*, Individual fibres of some direct flight muscles are stained (solid arrows), others unstained (open arrows). $\times 450$.

the underlying mesoderm fails to differentiate, whereas when the mesoderm is removed, the ectoderm differentiates autonomously²⁰⁻²². These earlier results, the present paper and recent studies of developmental genetics give a consistent picture. In the ectoderm, determination and development depend on genetic elements which are locally deployed in segmental and subsegmental compartments^{4,7,8,23}. By contrast, the mesodermal cells do not appear to be as precisely determined, their development being at least partly dictated by the neighbouring ectodermal cells. Thus the cellular and genetic mechanisms of control seem to differ for the two germ layers.

We thank Mark Bretscher and Pat Simpson for discussion, Paul Johnston for assistance and Gary Struhl for help with the manuscript.

Received 26 October; accepted 3 December 1981.

- Hadorn, E. *Brookhaven Symp. Biol.* **18**, 148-161 (1965).
- Garcia-Bellido, A., Ripoll, P. & Morata, G. *Nature new Biol.* **245**, 251-253 (1973).
- Lawrence, P. A. *J. Embryol. exp. Morph.* **30**, 681-699 (1973).
- Morata, G. & Lawrence, P. A. *Nature* **265**, 211-216 (1977).
- Chan, L.-N. & Gehring, W. *Proc. natn. Acad. Sci. U.S.A.* **68**, 2217-2221 (1971).
- Illmensee, K. *Symp. R. ent. Soc. Lond.* **8**, 76-96 (1976).
- Lewis, E. B. in *Role of Chromosomes in Development* (ed. Locke, M.) 231-252 (Academic, New York, 1964); *Nature* **276**, 565-570 (1978).
- Garcia-Bellido, A. *Ciba Fdn Symp.* **29**, 161-178 (1975).
- Ursprung, H., Conscience-Egli, M., Fox, D. J. & Walliman, J. *Proc. natn. Acad. Sci. U.S.A.* **69**, 2812-2813 (1972).
- Reed, C. T., Murphy, C. & Fristrom, D. *Wilhelm Roux's Arch. dev. Biol.* **178**, 285-302 (1975).
- Poddy, C. A. in *The Genetics and Biology of Drosophila* Vol. 2d (eds Ashburner, M. & Wright, T. R. F.) 407-441 (Academic, New York, 1980).
- Lawrence, P. A. *J. Embryol. exp. Morph.* **64**, 321-332 (1981).
- Bryant, P. J. *J. exp. Zool.* **193**, 49-78 (1975).
- Brower, D. L., Smith, R. J. & Wilcox, M. *J. Embryol. exp. Morph.* (in the press).
- Vaney, C. *Annls Univ. Lyon* **2**, 1-171 (1902).
- Schneiderman, H. A. in *Mechanisms of Cell Change* (eds Ebert, J. & Okada, T.) 243-272 (Wiley, New York, 1979).

- Perez, C. *Archs Zool. exp. gén.* **5**, 1-275 (1910).
- Brower, D. L., Lawrence, P. A. & Wilcox, M. *Dev. Biol.* **86**, 448-455 (1981).
- Hotta, Y. & Benzer, S. *Nature* **240**, 527-535 (1972).
- Bock, E. *Wilhelm Roux's Arch. dev. Biol.* **141**, 159-247 (1942).
- Haget, A. *Biol. Bull. Fr. Belg.* **87**, 123-217 (1953).
- Sauer, E. *Wilhelm Roux's Arch. dev. Biol.* **147**, 302-354 (1954).
- Lawrence, P. A. *Cell* **26**, 3-10 (1981).
- Lindsley, D. & Grell, E. H. *Genetic variations of Drosophila melanogaster* (Carnegie, Washington, 1968).
- Ephrussi, B. & Beadle, G. W. *Am. Nat.* **70**, 218-225 (1936).
- Miller, A. in *Biology of Drosophila* (ed. Demerec, M.) 420-534 (Wiley, New York, 1950).

Electrical stimulation of hindlimb increases neuronal cell death in chick embryo

Ronald W. Oppenheim

Neuroembryology Laboratory, Department of Mental Health, Anderson Hall, Dorothea Dix Hospital, Raleigh, North Carolina 27611, USA

Ramon Núñez

Biological Sciences Group, University of Connecticut, Storrs, Connecticut 06268, USA

The daily treatment of chick embryos with neuromuscular blocking agents during the period of massive naturally occurring death of spinal motoneurons (days 5-10) reduces the normal loss of these cells^{1,2}. Because both pre- and postsynaptic blocking agents have this effect³, the cause seems to be related to a reduction in synaptic and/or muscular activity. If this interpretation is correct, an experimentally induced increase in nerve or muscle activity (and/or an increased activation of acetylcholine receptors, AChR), during the normal cell death period should accelerate or enhance the loss of motoneurons. Consistent with this prediction, we have found that daily treatment with the nicotinic receptor agonist, carbachol, or with the anticholinesterase (anti-ChE) agent, eserine, leads to a significant enhancement of cell death of spinal motoneurons³. However, because in the same study other related agents (for example nicotine) failed to alter normal cell death, we believed that a different approach might provide a better test of our prediction. We now report that short-term continuous electrical stimulation of the hindlimb (the musculature and nerve trunks) initiated at any time between stages 31 (day 7) and stage 34 (day 8) results in a significant reduction in the number of healthy motoneurons and an increase in the number of pyknotic (degenerating) motoneurons in the lumbar lateral motor column (LMC). As the number of pyknotic cells in the dorsal root ganglia (DRG) is unaltered, this effect seems to be specific to motoneurons.

White Leghorn chick embryos were treated with curare, α -bungarotoxin (α -BTX) or saline as previously described². Briefly, beginning on day 5, 100-200 μ l of a curare (2-2.5 mg per embryo), α -BTX (75 μ g per embryo) or 0.9% saline solution was injected daily onto the vascularized chorioallantoic membrane (CAM). About half the embryos in each group (saline and curare or α -BTX) were used for electrical stimulation while half were subjected to sham procedures. The curare and α -BTX treatment was associated with a significant reduction in the rate of spontaneous neuromuscular activity (embryonic motility)^{1,2}. Data from the curare and α -BTX embryos were pooled (curare/ α -BTX). To gain access to the embryo's leg, a relatively large lateral window was made in the shell, and the CAM and amnion were cut, avoiding most blood vessels. In all cases, the right leg was used for both the experimental (stimulation) and sham procedures. Stimulation suction electrodes (2.5 mm tip diameter), made from polyethylene tubing, were filled with a saline solution and held in a micromanipulator for attachment to the thigh region of the leg. The leg of the embryo was stabilized by either pulling it out of the amniotic fluid and placing it on top of the CAM or by securing it with a hair loop which was attached to the shell. Despite these

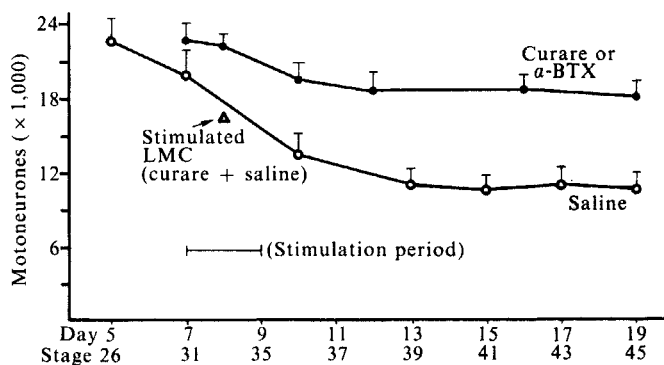


Fig. 1 The average number of motoneurons ($\pm$ s.d.) in the lumbar LMC of saline (O) and curare/ α -BTX (●) treated embryos. Embryos were treated as described by Pittman and Oppenheim². Δ , The average number of motoneurons in the right LMC of all experimental embryos in the present study (saline plus curare- α -BTX treated, stages 31–35) following electrical stimulation of the right hindlimb at various times during the stimulation period (—) (see text for details).

precautions, the suction electrode occasionally became detached from the leg. Consequently, the embryos were checked every 3–4 h and in those cases where the electrode had lost contact it was re-attached. Following the attachment of the electrode, the window in the egg was closed with Parafilm and the egg placed in a temperature-controlled environment of 33–35 °C. Electrical pulses were delivered via a Grass stimulator (duration 0.6 ms, rate 0.5–1.0 s⁻¹). The stimulus intensity was adjusted to 1.5–2.0 times the minimal intensity required to produce visible contractions of the leg. It was sometimes necessary to readjust the stimulus intensity during the experiment. The sham procedure was similar in all respects to that described above except that the stimulator was not turned on. At the end of the stimulation period, the surviving embryos (<12% of the total) were killed, staged⁴, fixed and processed for histological examination. The histological procedures and the methods for counting healthy and pyknotic neurones were as previously described^{5,6} (also see Fig. 2b legend).

From a total of 210 embryos, 24 survived and were killed after 5–22 h of treatment (mean 12.8 and 14.3 h for experimental and control respectively). Of these, nine (five experimental, four control) were found to have various degrees of limb necrosis caused either by the manipulation involved in attaching the electrodes or by electrolytic damage. Two animals in each group (experimental and control) had extensive necrosis of the leg and were excluded from further analysis. (Compared with the other experimental cases these embryos had more than double the number of pyknotic neurones in both the lumbar DRG and LMC.) The other five embryos (three experimental, two control) had only a small amount of necrosis, limited to the foot region, and thus are included in the data analysis. With these exceptions, based on a preliminary light microscopic analysis of limb musculature, there were no obvious differences in muscle histology between the experimental and control limbs. Most of the experimental embryos used for analysis were still showing visible limb contractions in response to the electrical stimulation at the end of the experiment.

Figure 1 shows an analysis of the changes in motoneurone number during development of the lumbar LMC of control saline and curare/ α -BTX treated embryos, together with the pooled data on healthy neurones from the electrically stimulated embryos. It can be seen that both stimulation periods (stages 31–33 and 34–35) fall within the time span of massive naturally occurring cell death.

Because no differences were found in the number of either healthy or pyknotic motoneurones between the right and left LMC of both the saline and curare/ α -BTX control (sham) embryos, the data from the two sides were pooled for all subsequent comparisons with the experimental, stimulated embryos. In all comparisons between the control and experi-

mental embryos, or between the stimulated and unstimulated side of the experimental embryos, there were always fewer healthy motoneurones in the right (stimulated) LMC of the experimental cases (Fig. 2a). In most of the comparisons these differences were statistically significant, although in the case of the stage 31–33 saline-treated group there were too few surviving embryos for a statistical comparison. However, when the data from all experimental embryos are pooled (saline + curare) the differences are statistically significant both at stages 31–33 and 34–35 ($P < 0.01$, Fig. 2a). In both groups there was a ~20% reduction in the number of neurones in the right (stimulated) LMC of the experimental embryos.

Comparison of the number of pyknotic cells in the LMC of control and experimental cases, or between the two sides of the spinal cord in the experimental cases showed that most of the differences were statistically significant. When the two sides of the spinal cord of the pooled experimental cases (saline + curare) were compared statistically, the differences were found to be highly significant ($P < 0.01$, Fig. 2b). At both stages there were approximately twice as many degenerating cells in the stimulated (right) LMC.

The possibility that prolonged electrical stimulation may somehow be lethal to developing neurones and that the effect observed is a pathological phenomenon unrelated to the mechanisms that normally regulate natural cell death seems unlikely for two reasons. First, it has been reported that even more prolonged periods of electrical stimulation of the chick embryo wing, after the normal period of extensive loss of brachial motoneurones⁷, does not induce cell loss⁸. Second, the DRG cells, which are also undergoing natural cell death during the stages examined in these experiments⁵, were unaffected by the electrical stimulation (Fig. 2b).

An additional procedural consideration concerns the rate and pattern of stimulation used here. The rate of stimulation

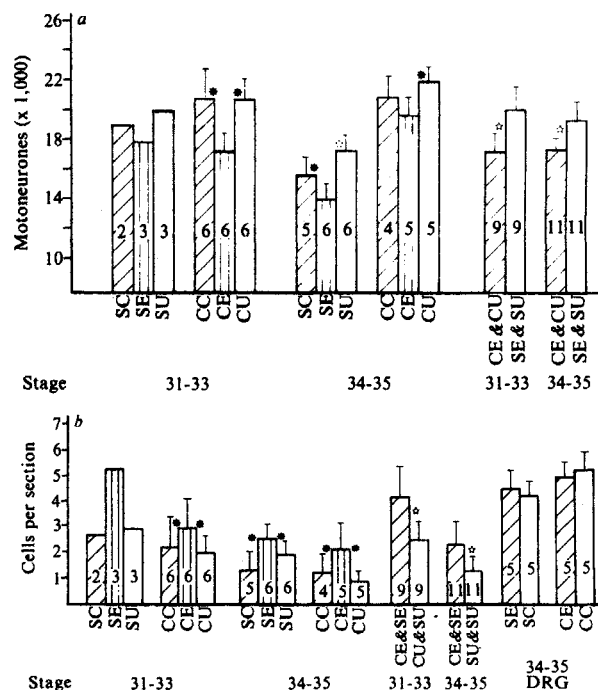


Fig. 2 The number of 'healthy' motoneurones (mean $\pm$ s.d.) (a) and of dying (pyknotic) motor and sensory neurones (mean $\pm$ s.d.) (b) in the lumbar LMC following electrical stimulation of the right hindlimb at stages 31–33 or 34–35. SC, saline control; SE, saline experimental (stimulated or right) LMC; SU, saline unstimulated (left) LMC; CC, curare/ α -BTX control; CE, curare/ α -BTX experimental (stimulated or right) LMC; CU, curare/ α -BTX unstimulated (left) LMC. CE & SE and CU & SU represent pooled data from the saline plus the curare/ α -BTX treated embryos. ★ $P < 0.05$; ☆ $P < 0.01$, Mann-Whitney tests, one-tailed. All statistical comparisons are relative to the experimental (stimulated) group, SE or CE. Sample size is as indicated in the bars.

(0.5–1.0 s⁻¹) was considerably higher than the normal rates of limb motility at these ages, which is between 6 and 12 per min (ref. 2). Although we chose a high rate of stimulation to increase our chances of finding an effect, we cannot rule out the possibility that the results reflect non-physiological conditions. However, in adult mammals a wide variety of stimulus rates and patterns are effective in maintaining the normal properties of denervated muscle⁹. We are now attempting to determine the effects of differing rates and patterns of stimulation on motoneurone death in the chick embryo.

We have suggested a model for explaining the role of activity in the regulation of neuronal cell death^{1,2,10}. Briefly, we proposed that during embryonic development, synaptic and/or muscle activity regulates those events which lead to a reduction in the availability of a trophic substance and/or in the level of AChR in muscle. This reduction would limit the number of synaptic sites that can be formed on a muscle cell, leading, in turn, to competition between neurones for a relatively fixed number of synaptic sites (and/or for a limited supply of trophic substance), in which competition the losers undergo cell

death. We suggest that neuromuscular blockade suppresses and electrical stimulation enhances this series of normal events.

Although many questions remain unanswered concerning the specific cellular and molecular events by which neuromuscular interactions regulate the number of motoneurones that survive the cell death period, the present findings provide additional support for our contention that synaptic and/or muscular activity is involved. In view of the increasing evidence that many other properties of the neuromuscular system depend on physiological activity for their normal differentiation and maintenance^{10–15}, it is perhaps not surprising that motoneurone cell death should now be added to this list.

This work was supported by NSF grant BNS 40397, by a grant from the Amyotrophic Lateral Sclerosis Society (ALS) and by funds provided by the North Carolina Department of Mental Health. Part of this work was done during the tenure of a Muscular Dystrophy Association Fellowship to R.N. We thank R. Jones for technical assistance, and Lanny Haverkamp and Jerome Maderdrut for their helpful comments.

Received 14 August; accepted 26 October 1981.

1. Pittman, R. & Oppenheim, R. W. *Nature* **271**, 364–366 (1978).
2. Pittman, R. & Oppenheim, R. W. *J. comp. Neurol.* **187**, 425–446 (1979).
3. Oppenheim, R. W. & Maderdrut, J. L. *Neurosci. Abstr.* **7**, 291 (1981).
4. Hamburger, V. & Hamilton, H. L. *J. Morph.* **88**, 49–92 (1951).
5. Hamburger, V., Brunso-Bechtold, J. L. & Yip, J. J. *Neurosci.* **1**, 60–71 (1981).
6. Chu-Wang, I. & Oppenheim, R. W. *J. comp. Neurol.* **177**, 33–58 (1978).
7. Oppenheim, R. W. & Majors-Willard, C. *Brain Res.* **154**, 148–152 (1978).
8. Toutant, M. & LeDourain, G. *IRCS med. Sci.* **8**, 408–409 (1980).
9. Lomo, T. & Westgaard, R. H. *J. Physiol., Lond.* **252**, 603–626 (1975).
10. Oppenheim, R. W. in *Studies in Developmental Neurobiology: Essays in Honor of Viktor Hamburger* (ed. Cowan, W. M.) 74–133 (Oxford, New York, 1981).
11. Brown, M. C., Holland, R. L. & Hopkins, W. G. A. *Rev. Neurosci.* **4**, 17–42 (1981).
12. Dennis, M. J. A. *Rev. Neurosci.* **4**, 43–68 (1981).
13. Fambrough, D. M. *Physiol. Rev.* **59**, 165–227 (1979).
14. Lomo, T. *Trends Neurosci.* **3**, 126–129 (1980).
15. Pestronk, A., Drachman, D. B., Stanley, E. F., Price, D. L. & Griffin, J. W. *Expl Neurol.* **70**, 690–696 (1980).

Significance of the potassium signal from neurones to glial cells

V. W. Pentreath & M. A. Kai-Kai

Department of Biology, University of Salford, Salford M5 4WT, UK

Neurones transmit signals to glial cells by releasing potassium into the intercellular spaces during nerve activity, simultaneously depolarizing their membrane potential¹. It is thought that the glial cells may respond metabolically or trophically to the released K, and that they may also take up excess K, thereby protecting the neurones against perturbations in the extracellular concentration of this ion^{2,3}. We have studied metabolic interactions between the two cell types in leech and snail ganglia using a recently discovered application^{4,5} of the radioactively labelled 2-deoxyglucose (2-DG) technique⁶. This approach derives from the finding that ³H-2-DG is partially and selectively metabolized into glycogen within nervous tissue, and thus can be used as a sensitive, specific marker for mapping sites of glycogen synthesis and distribution in nervous tissue^{4,5}. In the leech and snail ganglia, the large size of the neuronal perikarya, together with the grouping of the glial cells around them and the location of the axonal and dendritic processes in the core of the ganglia, allows the distributions of the labelled glycogen in the different cell types and their processes to be resolved by light and electron microscope autoradiography^{7,8}. Here we describe how in both animals, antidromic stimulation of nerves entering the ganglia increased the incorporation of ³H-2-DG into glycogen in glial cells surrounding the activated neurones. Increased extracellular K had a similar effect, which was maximal for a K concentration ~4 mM above normal. We therefore conclude that neuronal activation causes a K-mediated increase in ³H-2-DG uptake and synthesis into glycogen by the satellite glial cells.

In the technique used, incorporation of ³H-2-DG into 2-deoxyglycogen proceeds analogously to the glucosyl transfer from uridine diphosphate glucose (UDPG) into glycogen; the enzymes for glycogen synthesis are relatively nonspecific to the presence or absence of the hydroxyl group at C-2 of the glucose unit. The labelled 2-deoxyglycogen is able to withstand autoradiographic processing with aqueous media, whereas unmetabolized 2-DG and 2-DG-6-PO₄ are washed out. The resulting autoradiographs show the distribution of 2-DG

incorporated into glycogen, which corresponds to the amounts of endogenous glycogen.

Isolated abdominal ganglia of the leech, *Haemaphysalis sanguisuga*, and buccal ganglia of the snail, *Planorbis corneus*, were immersed in appropriate salines^{9,10} containing 5–10 μM 2-deoxy-D[1-³H]glucose (15.0 Ci mmol⁻¹; Radiochemical Centre, Amersham) and incubated at room temperature for 5–60 min. The effects of increased nerve activity were studied by stimulating nerves from the ganglia with suction electrodes, using pulses of 2.5–5.0 ms duration, 1–10 V at 5 Hz for between 10 min and 1 h. The uptake and metabolism of 2-DG in the presence of increased extracellular K were measured by incubating ganglia in saline solutions containing additions of 4 mM and 15 mM K at room temperature for 15 min to 1 h. Following the load periods with ³H-2-DG, all experimental tissues were analysed by three procedures. First, the levels of radioactivity in the ganglia were measured by liquid scintillation counting. Second, the amounts of ³H-2-DG incorporated into glycogen were assessed by the method of Mordoh *et al.*¹¹ (this method extracts high molecular weight glycogen in a pure form; the glycogen is not degraded and loss of label is minimized). Third, the distributions of labelled glycogen in the ganglia were determined by light and electron microscope autoradiography of tissues fixed with aldehyde and processed with aqueous media for histology⁴.

During 15–60 min exposure to ³H-2-DG, the ganglia of *Haemaphysalis* and *Planorbis* progressively accumulated ³H-2-DG at a rate of 0.13 and 0.16 nmol per g nervous tissue per min respectively. After 1 h, 4–5% of the ³H-2-DG was recovered in the high molecular weight glycogen fraction from *Haemaphysalis*, and ~1% in the fraction from *Planorbis* (Fig. 1b). The autoradiographic pattern of glycogen labelling observed by light microscopy differed in the two species. In *Haemaphysalis*, small areas of radioactivity were located in a narrow band of the peripheral cytoplasm of most neuronal perikarya, often in clumps of 1.5 μm diameter (Fig. 2a); axonal and dendritic processes were unlabelled. The giant glial cells^{7,8} surrounding the perikarya were labelled. In *Planorbis*, groups of neurones were selectively and heavily labelled after 1 h exposure to ³H-2-DG (Fig. 2c). Serial section analysis revealed that the radioactivity was distributed evenly throughout the axonal and dendritic processes. The glial cells surrounding the neurones were not labelled or only weakly. Electron microscope autoradiography showed that

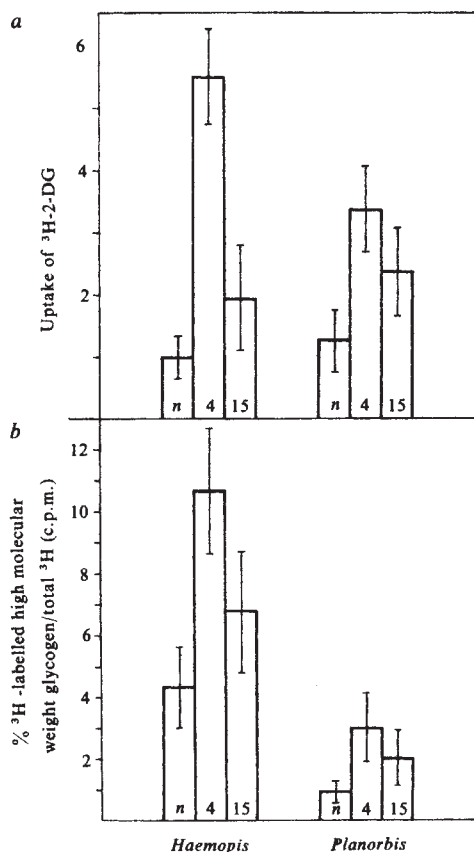


Fig. 1 Stimulation of ³H-2-DG uptake (a) and incorporation into high molecular weight glycogen (b) in abdominal ganglia of *Haemopsis* and buccal ganglia of *Planorbis*. Each set of histograms shows results for normal saline K concentrations, n (4.0 mM for *Haemopsis*; 1.3 mM for *Planorbis*), and for salines containing additional 4.0 mM and 15.0 mM K, added in the form of sulphate ions. Results are the means of three experiments in each case; bars indicate s.e. a, The ganglia were rinsed in saline (1 min) after the load period, homogenized in a tissue solubilizer (NCS; Radiochemical Centre, Amersham) and counted by liquid scintillation. All counts were corrected against blank controls (ganglia processed in parallel but without exposure to radioactive tracer) and for quenching (generally <1% using NCS solubilizer) by conventional methodology. The uptake of radioactivity by a unit weight of ganglia is expressed (in c.p.m.) relative to the uptake by *Haemopsis* ganglia in normal saline (unity). b, After the load period, high molecular weight undegraded glycogen was separated by a method similar to that described elsewhere¹¹. Ganglia were homogenized in ~3 vol 3% HgCl₂ and centrifuged at 2,000g for 5 min. The supernatant was added to ~2 vol of 95% ethanol, and the resulting precipitate separated by centrifugation (2,000g for 5 min). The precipitate was redissolved in water (0.5 ml) and the solution centrifuged again. The resulting supernatant was mixed with 2 vol of ethanol, and the sample was centrifuged at 2,000g for 10 min. The final sediment (the high molecular weight glycogen) was dissolved in NCS solubilizer and the radioactivity measured by liquid scintillation counting. Counts are expressed as percentages of the total radioactivity recovered in all the fractions. The lower molecular weight glycogen is not readily extracted by this procedure, thus the total percentages of labelled glycogen are likely to be higher.

the differences in labelling patterns were correlated directly with differences in the distribution of glycogen in the two species⁴. In *Haemopsis*, aggregates of glycogen particles were frequently contained within glial processes extending into the neuronal somata (Fig. 3).

Stimulation of nerves (1 h at 5 Hz) entering the ganglia of both species resulted in an approximate fivefold increase in accumulated radioactivity as measured by liquid scintillation. However, the incorporation of ³H-2-DG into glycogen was not proportional to this increase; the levels of labelled high molecular weight glycogen extracted from the tissues were increased 2.3–3.2 times (*Haemopsis*, mean increase = 3.2 ± 1.4 , 4 experiments; *Planorbis*, mean increase = 2.3 ± 1.1 , 4 experiments). When the distribution of labelled glycogen was studied by light microscope autoradiography, it was found that the labelling of the neurones, particularly their perikarya, was increased in both species, but a striking feature was the increased grain density in the glial cells

surrounding the perikarya (Fig. 2b, d). Thus neuronal activation caused an increase in uptake of ³H-2-DG by the nervous tissues, and to a lesser extent its incorporation into high molecular weight glycogen. As the neurones were stimulated via nerves outside the ganglia, the increased levels of glycogen in the glial cells inside the ganglia resulted indirectly from the neuronal activation.

There is evidence that extracellular [K] may normally be increased by 4 mM during maintained nerve activity in leech abdominal ganglia¹², and by 4–8 mM in the cat cortex¹³. Higher experimental levels of extracellular K produce increased depolarization of glial cells, but it is unlikely that such levels are of normal physiological significance³. The role of K in mediating the activity-induced increase in ³H-2-DG uptake and incorporation into glycogen was investigated by incubating ganglia in saline containing ³H-2-DG with elevated K levels and measuring the uptake of radioactivity and its conversion to glycogen. Other ganglia incubated in parallel were processed for autoradiography. Exposure of *Haemopsis* ganglia to a wide range of extracellular [K] from the normal level (4.0 mM) to 40 mM in 1 mM increments for 1 h revealed a maximal increase in ³H-2-DG uptake (about five times) when the increase was 3–5 mM (Fig. 1a). Higher values of extracellular K gave lower increases in 2-DG uptake, which varied between 1.5 and 2.2 times, although increases of up to four times were occasionally ob-

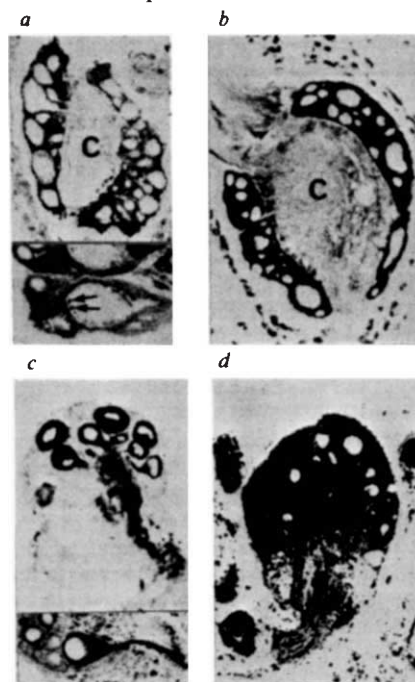


Fig. 2 Distribution of labelled glycogen in the abdominal ganglion of *Haemopsis* (a, b) and the buccal ganglion of *Planorbis* (c, d). The ganglia were exposed to ³H-2-DG for 1 h, then processed by aldehyde fixation, water-ethanol dehydration and wax sectioning⁴. The autoradiographs were prepared by coating the sections with Kodak AR 10, storing for 2 weeks, and then developing in Kodak D-19. a, Top: the labelled glial cells surround the neuronal perikarya which are located in the cortex of the ganglion. The core of the ganglia (C) contains large numbers of axonal and dendritic processes, but few glial cell processes⁷; bottom: at higher magnification grain clumps can be distinguished in the peripheral somatoplasm of some neurones (arrows). Many of these correspond to the glial invaginations visualized by electron microscope autoradiography (Fig. 3). b, Section of a ganglion stimulated antidromically via the abdominal connective at 5 Hz for 1 h in the presence of ³H-2-DG. Note the heavy labelling of the glial cells. c, Labelled neurones in a buccal ganglion of *Planorbis*: top, a group of labelled neurones which project into the neuropile of the ganglion; bottom, a large neurone cut in section to illustrate the equal distribution of labelled glycogen in the peripheral cytoplasm of the cell body and its axon. The branching patterns of the main axonal processes of the labelled neurones were mapped by comparison with adjacent serial sections. d, Section of buccal ganglion stimulated antidromically via the cerebro-buccal connective at 5 Hz for 1 h. Note the intense labelling of the glial cells which surround the neuronal perikarya, and increased labelling of the neurones. In both species the ganglia are avascular, but there are no restrictive barriers to the access of small molecules¹. All magnifications, $\times 100$, except for a, bottom ($\times 320$) and c, bottom ($\times 175$).

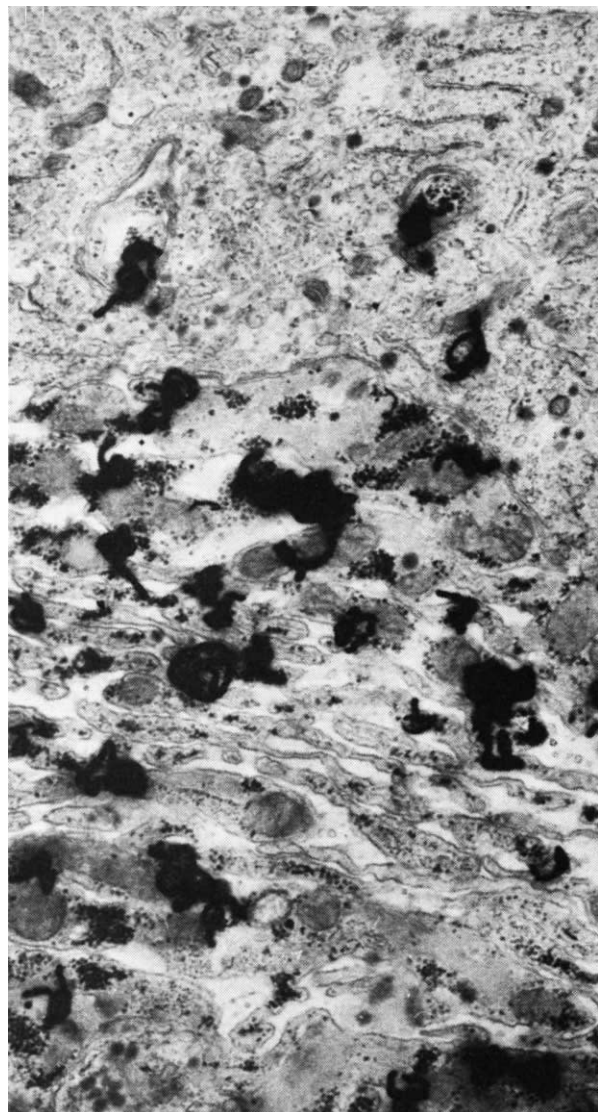


Fig. 3 Electron microscope autoradiograph of a section of the cortex of *Haemopsis* abdominalis ganglion. After exposure to ^3H -2-DG the tissue was preserved by fixation in glutaraldehyde and osmium solutions, and processed by aqueous histological methods⁴. Labelled glycogen deposits are located in the processes of a giant glial cell (bottom) adjacent to a neurone (top). Glycogen-filled glial processes invaginate the neurone ($\times 26,000$).

served at 25 mM. Similar findings were obtained for *Planorbis*. K concentrations of 4 and 15 mM above the normal saline values were used to study the effects of ^3H -2-DG incorporation into glycogen (Fig. 1b). The percentage of label extracted in the high molecular weight glycogen was increased 2.5-fold for *Haemopsis*, and 3-fold for *Planorbis* at 4 mM K, and to a lesser extent at 15 mM K. Autoradiographic analysis showed that 4 mM K (1 h exposure) produced labelling of glial cells that was comparable to the effects of electrical stimulation. A concentration of 15 mM caused some increase in the grain density overlying the glial cells, and also over the neurones. Extracellular recording experiments showed that the 4 mM increase in [K] caused no significant increases in the electrical activity of neurones in either species. Activity was initially greatly enhanced, but not maintained, at [K] > 12 mM.

These results provide direct evidence that increased neuronal activity causes a significant potassium-mediated increase in ^3H -2-DG uptake and incorporation into glycogen by the satellite glial cells. The similarity of the findings in both leech and snail ganglia indicate that this may be a general neurone-glial property. Both electrical stimulation and increased extracellular [K] increased the amounts of ^3H -glucose incorporated into glycogen in leech and snail ganglia (our unpublished obser-

vations), however the sites could not be resolved accurately by autoradiography as the label was also extensively incorporated into proteins and lipids.

The enzyme systems that synthesize and catabolize glycogen in mammalian brain are the same as those in other tissues^{14,15}, although they appear to differ in certain kinetic properties and modes of regulation¹⁶. It is evident that glycogen metabolism within the central nervous system must be controlled locally, because circulating hormones cannot pass the blood-brain barrier. There are no data on possible modulatory roles of K or K-induced depolarization on glycogen metabolism in nervous tissue. However, the selective stimulatory effects of K on glycogen formation in liver are well established^{17,18}. Glial cell membranes are selectively permeable to K⁺ and the K-mediated promotion of glycogen synthesis in glial cells caused by localized neuronal activity could provide a significant mechanism for ensuring continued energy supplies during any subsequent intense activity if local blood-glucose levels fell or were inadequate.

This work was supported by the SERC.

Received 3 July; accepted 9 November 1981.

1. Kuffler, S. W. & Nicholls, J. G. *Ergebn. Physiol.* **57**, 1-90 (1966).
2. Kuffler, S. W. & Nicholls, J. G. *From Neuron to Brain* (Sinauer, Sunderland, Massachusetts, 1976).
3. Varon, S. S. & Somjen, G. G. *Neurosci. Res. Prog. Bull.* **17**, (1) (1979).
4. Kai-Kai, M. A. & Pentreath, V. W. *J. Neurocytol.* **10**, 693-708 (1981).
5. Pentreath, V. W., Seal, L. H. & Kai-Kai, M. A. *Neuroscience* (in the press).
6. Sokoloff, L. *et al. J. Neurochem.* **28**, 897-916 (1977).
7. Kai-Kai, M. A. & Pentreath, V. W. *J. comp. Neurol.* (in the press).
8. Coggeshall, R. E. & Fawcett, D. W. *J. Neurophysiol.* **27**, 229-289 (1964).
9. Nicholls, J. G. & Kuffler, S. W. *J. Neurophysiol.* **27**, 645-671 (1964).
10. Berry, M. S. *J. exp. Biol.* **56**, 621-637 (1972).
11. Mordoh, J., Krisman, C. R. & Leloir, L. F. *Archs Biochem. Biophys.* **113**, 265-272 (1966).
12. Baylor, D. A. & Nicholls, J. G. *J. Physiol., Lond.* **203**, 555-569 (1969).
13. Somjen, G. A. *Rev. Physiol.* **37**, 163-190 (1975).
14. Coxon, R. V. in *Handbook of Neurochemistry* Vol. 3 (ed. Lajtha, A.) 37-52 (Plenum, New York/London, 1970).
15. Siesjö, B. K. *Brain Energy Metabolism* (Wiley, New York, 1978).
16. Goldberg, N. D. & O'Toole, A. G. *J. biol. Chem.* **244**, 3053-3061 (1969).
17. Buchanan, J. M., Hastings, A. B. & Nesbitt, F. B. *J. biol. Chem.* **180**, 435-445 (1949).
18. Hastings, A. B. & Buchanan, J. M. *Proc. natn. Acad. Sci. U.S.A.* **28**, 478-482 (1942).

Depletion of substance P-containing axons in substantia gelatinosa of patients with diminished pain sensitivity

John Pearson*, Leslie Brandeis* & A. Claudio Cuello†

* Department of Pathology, New York University Medical Center, New York, New York 10016, USA

† Departments of Pharmacology and Human Anatomy, University of Oxford, Oxford OX1 3QT, UK

Substance P is an undecapeptide which may be involved in transmission of nociceptive information at synapses of primary sensory neurones^{1,2}. It occurs in 10-20% of rat dorsal root ganglion neurones^{3,4}. Small substance P-immunoreactive (SP-IR) axons from dorsal root ganglia enter the spinal cord and form a dense terminal network in the substantia gelatinosa of the rat dorsal horn³; analogous fibres from the trigeminal ganglion terminate in the substantia gelatinosa of the spinal trigeminal tract of the medulla oblongata^{4,5}, while sensory axons travelling via the petrosal and nodose ganglia terminate in the tractus solitarius of the medulla⁵⁻⁸. Substance P is also present in human spinal cord and pallido-nigral system⁹⁻¹². In addition to sympathetic and parasympathetic abnormalities, patients with familial dysautonomia (the Riley-Day syndrome) have severe diminution of temperature and pain sensitivity, reduced populations of primary sensory neurones and small Lissauer's tracts¹³. It was thus postulated that a diminution of primary substance P axons might be demonstrable using immunocytochemistry. We report here the presence of anatomically discrete depletion of substance P immunoreactivity in the substantia gelatinosa of spinal cord and medulla of patients with familial dysautonomia. The results are consistent with the hypothesis that substance P is involved in the transmission of impulses related to pain.

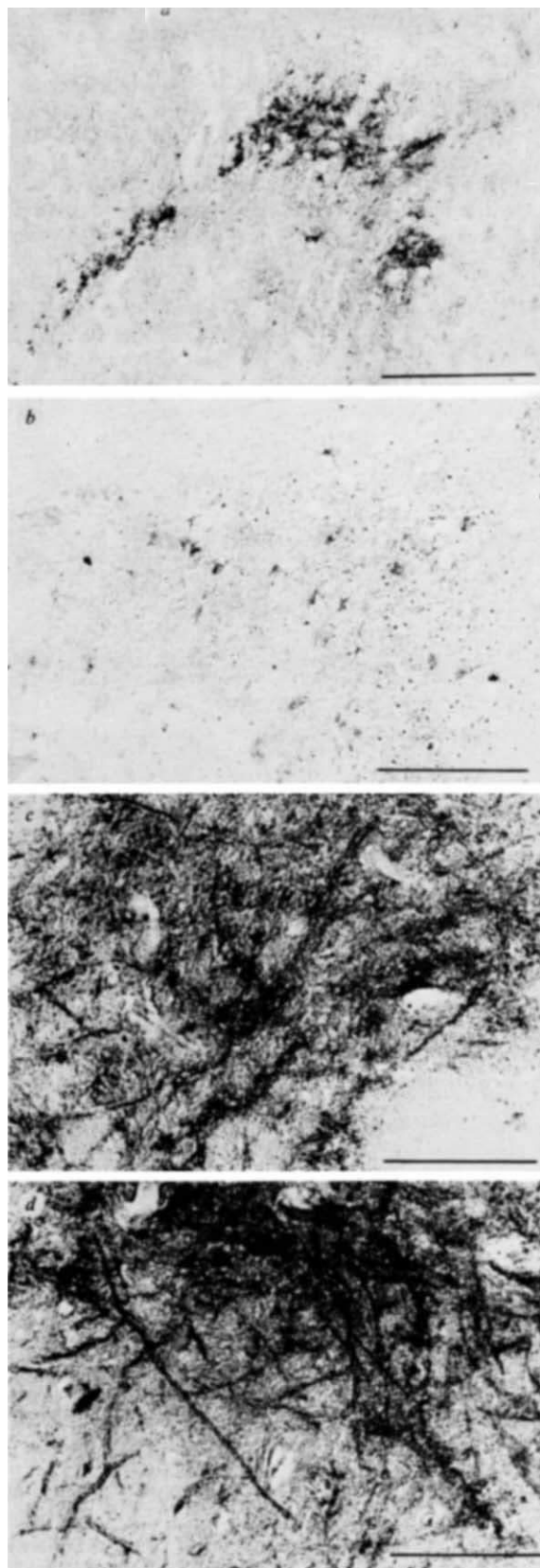


Fig. 1 The substantia gelatinosa of the control spinal cord (a) contains multiple substance P-immunoreactive (SP-IR) axons and their terminals, which form a distinct cap over the dorsal grey horn; no such cap is visible in the familial dysautonomia patient's spinal cord (b). The substantia nigra of both the control (c) and the patient (d) contain similar amounts of SP-IR fibres. Scale bars, 100 μ m.

The present studies used a well characterized monoclonal antibody (NCI/34HL) to substance P¹⁴. Thus, it does not cross-react with any other known mammalian neuroactive peptide, stains substance P in brain regions shown to contain the peptide by bioassay, and stains the peptide which accumulates proximal to peripheral nerve ligatures and which can be independently identified on Sephadex columns¹⁴. The antibody was used with the peroxidase-antiperoxidase (PAP) technique of Sternberger¹⁵. Staining controls included normal serum and antibody adsorbed with substance P, neither of which gave any positive reaction. Test tissues used were from five patients aged 14, 15, 17, 18 and 22 yr, diagnosed as having Riley-Day syndrome at the Dysautonomia Treatment and Evaluation Center of New York University Medical Center. The diagnosis was confirmed at autopsy by a characteristic combination of defects in the tongue, sympathetic ganglia and dorsal root ganglia. Spinal cord and brain stem sections were made from archival blocks of paraffin-embedded tissue which had been fixed in 10% formalin. Storage times for patient and control tissues were similar.

In contrast to the controls, which showed large concentrations of SP-IR terminals in the substantia gelatinosa of spinal cord and medulla oblongata (Fig. 1a), there was almost complete absence of such terminals from these regions in all patients with familial dysautonomia (Fig. 1b). Some of the few remaining SP-IR axons observed may have been derived from the brain stem, as studies in other species^{1,2} have shown the presence of SP-IR axons in this region. Patients did not differ from controls in the density of SP-IR axons in the substantia nigra (Fig. 1c,d). Immunoreactive axons were also readily demonstrable in the tractus solitarius of patients. It is not possible to obtain good staining of the small quantities of substance P normally present in the perikarya of primary sensory neurones of monkey or man, and colchicine pretreatment is necessary to cause sufficient accumulation of the peptide in most lower animals. We were therefore unable to demonstrate directly depletion of primary sensory substance P-containing neurones in the patients. Nevertheless, we have previously demonstrated severe losses of sensory axons, especially those of small diameter, in familial dysautonomia and have shown that severe deficits of dorsal root ganglion neurones occur in young patients¹³. Depletion of spinal dorsal horn neurones has not been demonstrated, and the fact that spinal dorsal column fibres only diminish as sensory neuronal loss continues into adulthood¹³ indicates that initial deficits are mainly among neurones projecting onto the substantia gelatinosa and other parts of the dorsal grey. Small-calibre sensory axons projecting onto the substantia gelatinosa and containing substance P have been suggested as being involved in processing of pain-related impulses^{1,2}. The coincidence of analgesia, loss of sensory neurones and small diameter axons projecting onto the substantia gelatinosa, and severe diminution of dorsal horn substance P immunoreactivity in familial dysautonomia support this idea. However, it does not prove the case because other sensory modalities, notably those related to temperature, are also attenuated in familial dysautonomia; other putative neurotransmitters yet to be examined may be found to be decreased in the disease.

The persistence of some substance P in axons of the tractus solitarius might be attributed to sources within the central nervous system. Alternatively, there might be less sensory neuronal loss in cranial ganglia, some of which are of placodal origin, than in dorsal root ganglia, which are of neural crest origin. No formal neuronal counts have been made of patients' cranial ganglia.

Although immunocytochemistry cannot definitively demonstrate the absence of an antigen, there is no doubt that the amount of substantia gelatinosa axon substance P in the patients was far below normal. That this was not an artefact of autolysis is indicated by comparison with controls in which the postmortem intervals were generally longer than those of the patients but in which substantia gelatinosa SP-IR axons were readily demonstrable. Furthermore, substance P has been demonstrated to

have excellent postmortem stability^{12,16,17}. Finally, the ready detection of SP-IR axons in the tractus solitarius and substantia nigra of the patients provides excellent internal control both of staining efficacy and postmortem preservation of the peptide. The fact that SP-IR is normal in these structures indicates that familial dysautonomia does not involve a generalized defect in substance P synthesis.

Our data indicate a need for prospective studies of substance P levels in fresh wet tissue from patients. They also demonstrate the usefulness of immunocytochemistry in the microscopic study of human neurobiology and in the detection of deficits in chemically defined neurones in human disease.

This work was supported by the Dysautonomia Foundation (to J.P., L.B.) and the MRC, Wellcome Trust and the Royal Society (to A.C.C.).

Received 4 August; accepted 6 November 1981.

1. Nicoll, R. A., Schenker, C. & Leeman, S. E. *A. Rev. Neurosci.* **3**, 227 (1980).
2. Henry, J. L. *Brain Res.* **114**, 439 (1976).
3. Hokfelt, T. *et al. Neuroscience* **1**, 131 (1976).
4. Cuello, A. C., DelFiacco, M. & Paxinos, G. *Brain Res.* **152**, 499 (1978).
5. Sumal, K. K., Pickel, F. M., Miller, R. J. & Reis, D. J. *Soc. Neurosci.* **6**, 353 (1980).
6. Maley, B. E. & Elde, R. P. *Soc. Neurosci.* **6**, 355 (1980).
7. Helke, C. J., O'Donohue, T. L. & Jacobowitz, D. M. *Peptides* **1**, 1 (1980).
8. Gillis, R. A., Helke, C. J., Hamilton, B. L., Norman, W. P. & Jacobowitz, D. M. *Brain Res.* **181**, 476 (1980).
9. Cuello, A. C., Polak, F. & Pearse, A. G. E. *Lancet* **ii**, 1054 (1976).
10. de Lanerolle, N. C. & LaMotte, C. *Soc. Neurosci.* **6**, 353 (1980).
11. Defendini, R., Nilaver, G., Zimmerman, E. A. & Tennyson, V. M. *Soc. Neurosci.* **6**, 103 (1980).
12. Kanazawa, I., Bird, E., O'Connell, R. & Powell, D. *Brain Res.* **120**, 387 (1977).
13. Pearson, J. J. *autonomic nerv. Syst.* **1**, 119 (1979).
14. Cuello, A. C., Galfre, G. & Milstein, C. in *Receptors for Neurotransmitters and Peptide Hormones* (eds Pepeu, G., Kuhar, M. J. & Enna, S. J.) 349 (Raven, New York, 1980).
15. Sternberger, L. *Immunocytochemistry* (Wiley, New York, 1979).
16. Cooper, P. E., Fernstrom, M. H., Leeman, S. E. & Martin, J. B. *Soc. Neurosci.* **6**, 680 (1980).
17. Emson, P. C., Arregui, A., Clement-Jones, V., Sandberg, B. E. B. & Rossor, M. *Brain Res.* **199**, 147 (1980).

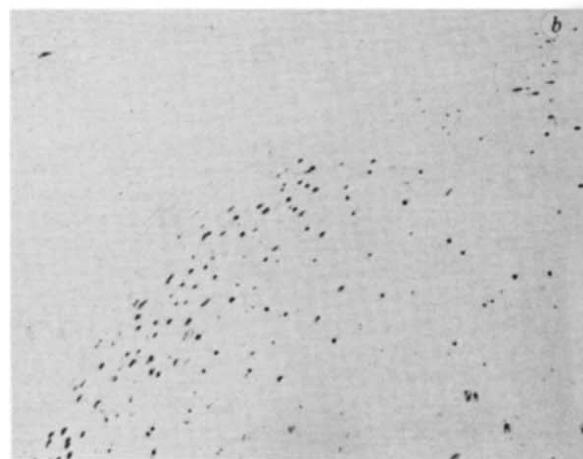
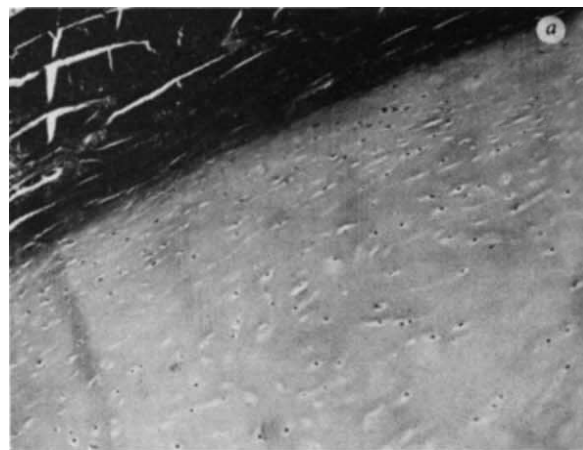


Fig. 1 Serial paraffin sections from the xiphoid process of human adult: *a*, haematoxylin and eosin stain ($\times 90$); *b*, immunoperoxidase reaction for S-100 protein, without counterstain ($\times 90$). Reaction product is present only within chondrocytes; the extracellular matrix of the cartilage and the surrounding fibrous connective tissue are unstained. *c*, Same section as *b* ($\times 500$).

S-100 protein in human chondrocytes

K. Stefansson*, R. L. Wollmann*, B. W. Moore†
& B. G. W. Arnason*

* Departments of Pathology and Neurology, University of Chicago, 950 East 59th Street, Chicago, Illinois 60637, USA

† Department of Psychiatry, Washington University School of Medicine, St Louis, Missouri 63110, USA

Chondrocytes in the fetal skull and visceral cartilage are considered to be of neural crest origin¹ while those found elsewhere in the body are thought to be derived from mesoderm^{2,3} (mesodermal chondrocytes). Chondrocytes of pelvic rudiments from chick embryos are the only cells known to respond to nerve growth factor that are not of neuroectodermal origin⁴. In addition, embryonic spinal cord⁵ and neural retina⁶ of chickens synthesize type II collagen^{5,6}, originally thought to be confined to cartilage. Here we report the presence of the S-100 protein (S-100) in chondrocytes of skull, vertebrae, ribs, sternum, larynx and long bones of the human fetus, and larynx and xiphoid process of the human adult. Mesodermal chondrocytes are the only cells which contain S-100 and yet are not considered to be of neuroectodermal origin, thus our results indicate that these chondrocytes may be developmentally closer to neuroectoderm than traditionally thought.

S-100, an acidic protein which has been isolated from the central nervous system (CNS) of mammals, has a molecular weight (M_r) of 21,000–24,000 (refs 7–9) and exists in both membrane-bound and soluble forms¹⁰. It has been shown to increase activity of RNA polymerase in nuclei isolated from brains of immature chickens¹¹. S-100 facilitates passage of γ -aminobutyric acid across nerve-cell membranes¹² and undergoes orthodromic axonal transport¹³. In the CNS the concentration of S-100 is low at birth, increasing with age^{7,14}.

S-100 has been detected by immunological methods in all vertebrates examined⁷. In the CNS, it is located primarily within

astrocytes and oligodendrocytes^{15–17}. Some investigators have also demonstrated S-100 in CNS neurones^{18–20} and in isolated synaptosomes¹⁰. Outside the CNS, S-100 has so far been demonstrated only in Schwann cells of peripheral nerve and in satellite cells of dorsal root and autonomic ganglia²¹.

While looking for S-100 in the fetal nervous system we discovered that the protein is present in chondrocytes of the vertebral column and decided to determine its distribution using anti-S-100 antiserum and the peroxidase-antiperoxidase method described elsewhere²². Rabbit anti-S-100 antiserum was raised by five or more intradermal injections of bovine

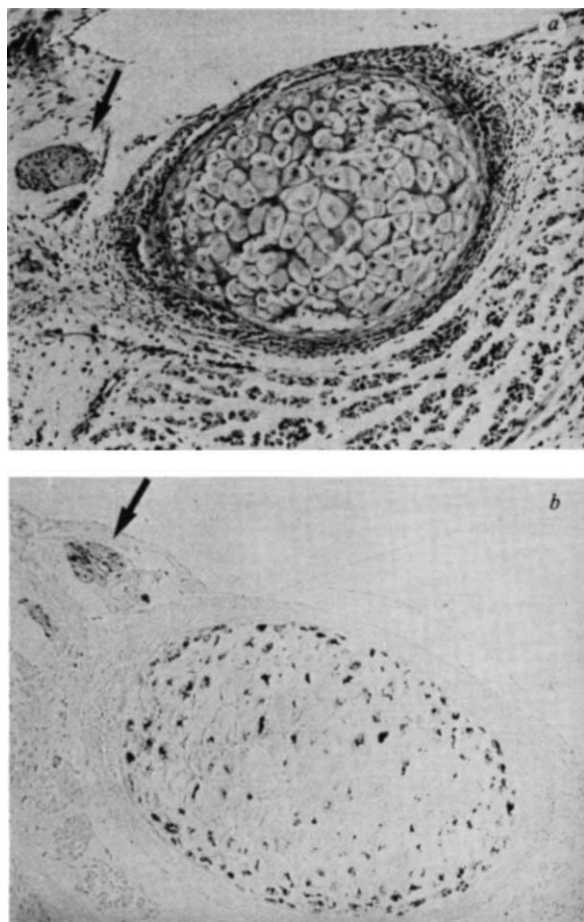


Fig. 2 Serial paraffin sections from a 7-week-old human embryo. *a*, Haematoxylin-eosin stain of a rib and adjacent intercostal nerve (arrow), $\times 100$. *b*, Immunoperoxidase reaction for S-100 protein, $\times 100$. Reaction product is present in chondrocytes of rib and Schwann cells of intercostal nerve. The adjacent fibrous connective tissue and skeletal muscle are unstained.

S-100 (isolated as described previously^{7,23}), mixed with complete Freund's adjuvant. As control sera we used anti-S-100 antiserum absorbed with S-100 until it no longer reacted with human brain, and serum from a rabbit which had been immunized with complete Freund's adjuvant alone. The anti-S-100 antiserum and the control sera were used at dilutions containing the same concentration of immunoglobulins. Tissues studied were from human fetuses of various ages and included cranium, larynx, trachea, vertebrae, ribs and lower and upper extremities. Xiphoid processes and laryngeal cartilage of adults were also examined. All tissues were fixed in formalin, embedded in paraffin and cut into 6- μ m sections.

Chondrocytes in all tissues examined were stained with anti-S-100 antiserum. Staining was both cytoplasmic and nuclear; the extracellular matrix was unstained. Control sera did not stain any of the tissues (Figs 1, 2).

Chondrocytes which form the fetal skull as well as the visceral cartilage are said to be of neural crest origin¹ and those found elsewhere are considered to be derived from mesoderm^{2,3}. Nevertheless, it is accepted that tissues of neuroectodermal origin can exert a considerable influence on differentiation of mesodermal chondrocytes³. Extirpation of the neural tube from chick embryos inhibits formation of axial cartilage, and grafting of spinal cord with or without notochord and neural crest into any region of the somite of tail-bud embryos results in formation of a complete secondary vertebral column². The standard interpretation of these phenomena has been that the spinal cord induces formation of cartilage in the somites³.

Type II collagen was originally thought to be unique to cartilage but recently it has been demonstrated that neural

retina⁶ and probably spinal cord⁵ of chick embryos also produce type II collagen. In addition, nerve growth factor inhibits chondromucoprotein synthesis by cartilage of pelvic rudiments from chick embryos⁴. Mesodermal chondrocytes are the only cells known to respond to nerve growth factor but not thought to be of neuroectodermal origin.

We have shown that S-100 is present in both mesodermal and neural crest chondrocytes of man. The presence of S-100 in chondrocytes of skull, larynx and trachea is perhaps not unexpected, but the fact that we have found S-100 in chondrocytes of vertebrae, ribs, xiphoid processes and extremities suggests that mesodermal chondrocytes may be developmentally closer to neuroectoderm than was previously believed.

Note added in proof: The S-100 protein has recently also been identified in melanocytes and Langerhans cells of normal human skin by D. Cocchia, F. Michetti and R. Donato (*Nature* **294**, 85-87 (1981)).

Received 20 August; accepted 5 November 1981.

1. LeLivre, C. *J. embryol. exp. Morph.* **31**, 453-477 (1974).
2. Holzer, H. & Detwiler, S. R. *J. exp. Zool.* **123**, 335-369 (1953).
3. Hall, B. K. *Adv. Anat. Embryol. Cell Biol.* **53**, 1-49 (1978).
4. Eisenbarth, G. S., Drezner, M. K. & Lebovitz, H. E. *J. Pharmac. exp. Ther.* **192**, 630-634 (1975).
5. Trelstad, R. L., Kang, A. H., Cohen, A. M. & Hay, E. D. *Science* **179**, 295-297 (1973).
6. Smith, G. N., Leisenmayer, T. F. & Newsome, D. A. *Proc. natn. Acad. Sci. U.S.A.* **73**, 4420-4423 (1976).
7. Moore, B. W. & Perez, V. J. in *Physiological and Biochemical Aspects of Nervous Integration* (ed. Carlson, F. D.) 343-368 (Prentice-Hall, Englewood Cliffs, New Jersey, 1968).
8. Moore, B. W. *Biochem. biophys. Res. Commun.* **19**, 739-744 (1965); *Int. Rev. Neurobiol.* **15**, 215-255 (1972).
9. Dannies, P. S. & Levine, L. *J. biol. Chem.* **246**, 6276-6283 (1971).
10. Donato, R. & Michetti, F. *Experientia* **30**, 511-512 (1974).
11. Miani, N., Michetti, F., DeRenzi, G. & Ganiglia, A. *Experientia* **29**, 1499-1501 (1973).
12. Hyden, H., Lange, P. W. & Larson, S. *J. neurol. Sci.* **45**, 303-316 (1980).
13. Miani, N., DeRenzi, G., Michetti, F., Sangiacomo, C. O. & Ganiglia, A. *J. Neurochem.* **19**, 1387-1394 (1972).
14. Zuckerman, J. E., Herschmann, H. R. & Levine, L. *J. Neurochem.* **17**, 247-251 (1970).
15. Perez, V. J., Olney, J. W., Cicero, T. J., Moore, B. W. & Bahn, B. A. *J. Neurochem.* **17**, 511-519 (1970).
16. Ludwin, S. K., Kosek, J. C. & Eng, L. F. *J. comp. Neurol.* **165**, 197-208 (1976).
17. Matus, A. & Mughals, S. *Nature* **258**, 746-748 (1975).
18. Sviridou, S. M., Korochkin, L. I., Ivanov, V. N., Maletskaya, E. I. & Bakhtina, T. K. *J. Neurochem.* **19**, 713-718 (1972).
19. Hansson, H. A., Persson, L., Rönnebäck, L. & Haglid, K. G. *Cytobios* **15**, 45-48 (1976).
20. Hyden, H. & McEwen, B. S. *Proc. natn. Acad. Sci. U.S.A.* **55**, 354-358 (1966).
21. Stefansson, K., Wollman, R. L. & Moore, B. W. *Brain Res.* (in the press).
22. Stefansson, K. & Wollmann, R. L. *Acta neuropath.* **52**, 135-140 (1980).
23. Levine, L. & Moore, B. W. *Neurosci. Res. Prog. Bull.* **3**, 18-22 (1965).

A very small porcine virus with circular single-stranded DNA

I. Tischer, H. Gelderblom, W. Vettermann & M. A. Koch

Robert Koch-Institute of the Federal Health Office, Nordufer 20, 1000 Berlin 65, FRG

It was reported previously that cultures of the porcine PK-15 cell line (ATCC-CCL31) were chronically infected with a small virus, supposedly containing RNA¹. No gross cytopathic effect was seen in these cultures. During the search for an *in vitro* model of persistent virus infections the agent was studied in more detail. These studies showed the virus to have a diameter of 17 nm, and to contain a covalently closed circular single-stranded DNA with a molecular weight of 0.58×10^6 and a main capsid polypeptide with a molecular weight of 36,000. Of the animals tested, only pigs were found to have antibodies. On the basis of these properties the virus appears to be a member of a family of animal viruses so far not encountered. We have named the virus porcine circovirus (PCV).

The virus was obtained from PK-15 cells. It can be propagated in selected PK-15 clones which are free of porcine parvovirus, porcine papovavirus, mycoplasmas and of PCV. Virus was concentrated by ultracentrifugation of Freon-treated cells and supernatants and then purified by two cycles of CsCl-gradient centrifugation. Sucrose gradient centrifugation yielded a sedimentation coefficient for the virion of 52S. On isopycnic centrifugation in CsCl virus particles banded exclusively at a density of 1.37 g cm^{-3} . Electron micrographs (Fig. 1) confirmed that the

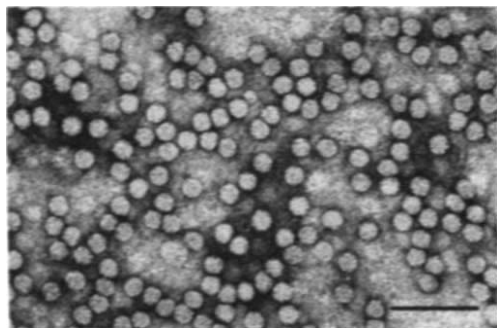


Fig. 1 Electron micrograph of purified PCV particles negatively stained with 1% aqueous uranyl acetate. Scale bar, 100 nm.

virus is an isometric particle with a diameter of 17 ± 1.3 nm. No stain-filled particles were observed.

Purified ^3H -leucine-labelled virus was disrupted at 100°C for 1 min with SDS (1%) and mercaptoethanol and the protein moiety was analysed by polyacrylamide gel electrophoresis². Only a single band containing radioactivity was found. Its position corresponded to that of a polypeptide of molecular weight $36,000 \pm 2,300$, using bovine serum albumin, ovalbumin and chymotrypsinogen as references.

For isolation and characterization of the nucleic acid, [^3H]thymidine-labelled virus was lysed with SDS (0.5%) and sodium *N*-lauroyl sarcosinate (1%) for 15 min at room temperature, followed by a digestion with proteinase K ($500 \mu\text{g ml}^{-1}$) for 4 h at 37°C . Nucleic acid was purified by centrifugation in a neutral sucrose gradient (5–20% wt/wt; rotor SW 40; 33,000 r.p.m.: 7.5 h; 20°C), and by precipitation

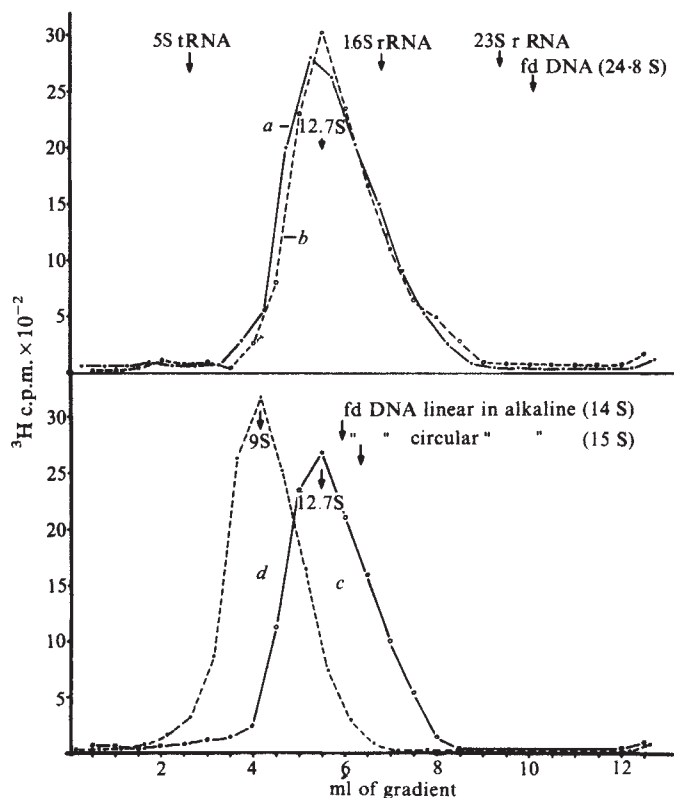


Fig. 2 Sedimentation profiles of [^3H]thymidine-labelled native, heated and alkali treated PCV DNA in sucrose gradients (5–20% wt/wt). a, Native DNA; b, heated DNA (10 min; 100°C); c, alkali treated (0.1 M NaOH; 5 min; 22°C) and subsequently neutralized (with 0.1 M HCl) DNA in neutral gradients (1 M NaCl; 0.01 M Tris; 0.001 M EDTA; pH 7.2); d, alkali-treated DNA in an alkaline gradient (0.9 M NaCl; 0.1 M NaOH; 0.001 M EDTA; pH 12.7). Beckman rotor SW 40; 33,000 r.p.m.; 7.5 h; 20°C . The sedimentation coefficients of PCV DNA were determined on the basis of the sedimentation of the markers (RNA) in the neutral gradient. For comparison virion fd-DNA was included.

with two parts of ice cold ethanol containing 0.4 M lithium chloride. The pellet was washed with 75% ethanol and then resuspended in $0.1 \times \text{SSC}$. The incorporation of ^3H via [^3H]thymidine, the buoyant densities and the alkali stability demonstrated that the nucleic acid of PCV is DNA. The buoyant densities were found to be 1.718 g cm^{-3} in CsCl using adeno 2 ^{14}C -DNA (1.715 g cm^{-3}) and DNA from *Micrococcus lysodeikticus* (1.731 g cm^{-3}) as markers, and in Cs_2SO_4 to be 1.45 g cm^{-3} . Purified tritium-labelled PCV DNA was kept in 0.3 M NaOH for 20 h at room temperature³. The radioactivity in trichloroacetic acid precipitates of treated and untreated nucleic acids was then assayed. While an untreated control yielded of the order of 5,500 c.p.m. the treated sample contained 5,750 c.p.m. The efficacy of this treatment on RNA was tested using 16S and 23S ribosomal RNA.

In transfection experiments⁴ native PCV DNA purified as above was shown to be infective ($\sim 0.04 \mu\text{g}$ PCV DNA per 1.5×10^6 cells). This infectivity was abolished completely by DNase 1—that is, using DNase-treated PCV DNA for transfection, cells producing virus, as revealed by staining with fluorescein-labelled specific antibodies, were not detectable even after two passages of the cells. In control experiments with untreated PCV DNA, more than 1,000 specifically stained cells per 1.5-cm diameter coverslips with about 2.5×10^5 cells were seen.

The sedimentation behaviour of native PCV DNA and of heat- or alkali-treated PCV DNA in neutral sucrose gradients as well as in alkaline gradients (Fig. 2) revealed that the DNA is single stranded⁵. The degradation of PCV DNA by the single-strand-specific nuclease S_1 confirmed this conclusion (Fig. 3). The sedimentation coefficient of PCV DNA determined in 1 M NaCl at pH 7.2 was 12.7S, and about 9S at pH 12.7. In 1.5% agarose gels the PCV DNA migrated about 2.3 times further than ΦX174 virion DNA.

When PCV DNA was examined by electron microscopy more than 85% of the molecules seen were circular (Fig. 4). To determine the size of the circular DNA, contour length measurements were performed using ΦX174 -am-3 virion DNA⁶ as an internal marker. From the data for ΦX174

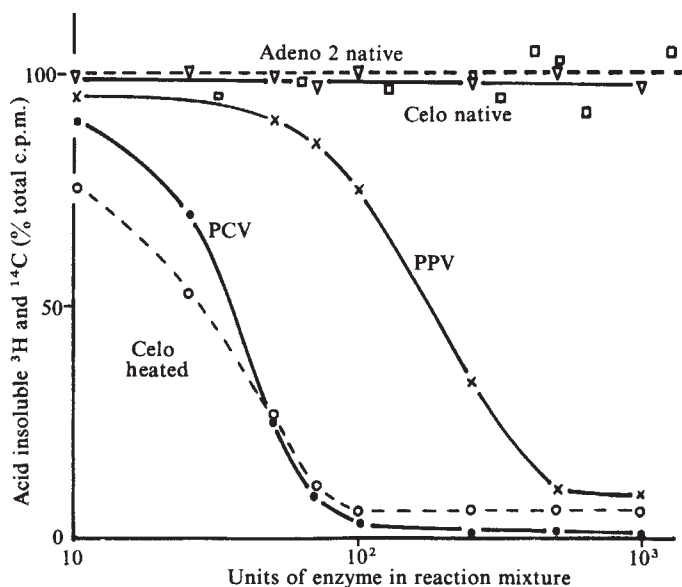


Fig. 3 Degradation of PCV DNA by single-strand-specific nuclease S_1 . PCV DNA was treated with nuclease S_1 (buffer: 0.1 M Na-acetate; 0.1 mM ZnSO_4 ; pH 4.5) in parallel with the double-stranded ^{14}C -adeno 2-DNA and ^3H -Celo-DNA as well as with the single-stranded ^3H -DNA from porcine parvovirus (PPV) and denatured (95°C , 10 min) ^3H -Celo DNA. DNA was incubated with nuclease S_1 for 30 min at 37°C and DNA was precipitated with trichloroacetic acid at a final concentration of 5% in the presence of $100 \mu\text{g}$ carrier RNA. Degradation is expressed as the relative loss of acid-precipitable radioactivity by increasing amounts of enzyme.



Fig. 4 Electron micrograph of PCV DNA (top) and Φ X174 virion DNA (bottom). A mixture of both species was spread from a hyperphase (carbonate buffer pH 10, 0.01% cytochrome c, 40% formamide) at room temperature. The specimen was rotary shadowed with Pt/Pd. Scale bar, 1 μ m.

($n = 198$, $1.94 \pm 0.12 \mu\text{m}$) and for PCV ($n = 212$, $0.633 \pm 0.05 \mu\text{m}$) a genome size of 1.76 kilobases, corresponding to a molecular weight of 0.58×10^6 , was calculated for PCV.

The sedimentation behaviour described above (Fig. 2) shows that the configuration of the molecule is unaffected by denaturing conditions, that is, after renaturation molecules of multiple unit length were never detected. This indicates that PCV DNA is a covalently closed circular molecule rather than a closed circle due to self-complementing cohesive ends. This interpretation is supported by the fact that no linearization of PCV DNA was detected when DNA was spread under heat denaturation. As internal controls *Hind*II, *Hind*III restriction fragments of fd-DNA carrying the transposon Tn5 (ref. 7) were included. These

restriction fragments consist of DNA molecules which are circular due to self-complementary ends of 190 and 340 bases respectively which readily opened in the conditions used.

Purified virus was used to raise specific antibodies in rabbits. Such antibodies aggregated PCV, neutralized infectivity and, using indirect immunofluorescence, detected virus producing cells. It was found that in chronically infected cultures only a small, but varying proportion of cells expressed virus antigens. PCV antisera did not inhibit agglutination of human red blood cells by porcine parvovirus. Attempts to demonstrate agglutination by PCV or RBC from man, rhesus monkey, pig, rabbit, guinea pig, mouse and chicken were unsuccessful.

Most probably the agent has its origin in pigs. A serological survey showed that 20 out of 22 randomly collected pig sera contained specific antibodies. Sera from rabbits, mice, calves and man, including sera from laboratory staff working with the agent, were negative. The physicochemical properties of this agent set it apart from all known animal viruses⁸. Similarities exist only with several plant pathogens comprising the group of geminiviruses^{9,10}.

We thank Renate Rasch, Doris Peters, S. Pociuli, H. Reupke and D. Elsen for technical assistance. Helpful suggestions were made by Professor H. Ludwig, Professor H. Köhler, and Drs Giovanna Morelli, G. Pauli, E. Werner and especially by Professor H. Schaller who also donated the fd derivative fd Km-30 and its restriction fragments.

Received 11 September; accepted 12 November 1981.

1. Tischer, I., Rasch, R. & Tochtermann, G. *Zbl. Bakt. Hyg., I. Abt. Orig. A* **226**, 153–167 (1974).
2. Laemmli, U.K. *Nature* **227**, 680–685 (1970).
3. Munro, H. N. *Meth. biochem. Anal.* **14**, 113–176 (1966).
4. Graham, F. L. & Van der Eb, A. J. *Virology* **52**, 456–467 (1973).
5. Studier, F. W. *J. molec. Biol.* **11**, 373–390 (1965).

6. Sanger, F. *et al. Nature* **265**, 687–696 (1977).
7. Herrmann, R., Neugebauer, K., Zentgraf, H. & Schaller, H. *Molec. gen. Genet.* **159**, 171–178 (1978).
8. *Classification and Nomenclature of Viruses: 3rd Rep. int. Committee on Taxonomy of Viruses* (ed. Matthews, R. E. F.) (Karger, Basle, 1979).
9. Goodman, R. M. *Virology* **83**, 171–179 (1977).
10. Harrison, B. D., Barker, H., Bock, K. R., Guthrie, E. J. & Meredith, G. *Nature* **270**, 760–762 (1977).

A monoclonal antibody against human T suppressor lymphocytes binds specifically to the surface of cultured oligodendrocytes

Joel Oger

INSERM U-49, 35043 Rennes, France

Sara Szuchet, Jack Antel & Barry G. W. Arnason

Department of Neurology and the Brain Research Institute, University of Chicago, Chicago, Illinois 60637, USA

The cardinal pathological feature of multiple sclerosis (MS) is the presence within brain and spinal cord of regions called plaques, from which myelin has been lost¹. It has been proposed that autoimmunity may be involved in MS and that oligodendrocytes, the cells responsible for the formation and maintenance of central nervous system myelin, may provide a target for immune attack in this disease. The fact that oligodendrocytes are reduced in number in MS plaques² is consistent with this hypothesis. MS is characterized by attacks and remissions. During attacks, a subset of T lymphocytes, known as T suppressor cells, is depleted from the circulation^{3–5}. The reason for this depletion is unknown, but the suppressor cells could be destroyed, sequestered, for example in brain, or modulated so that they are no longer detectable by the assays which enumerate them or which measure their functional capacities. Were suppressor cells and oligodendrocytes to share a determinant relevant to MS, both cell types could be targets for immune-mediated killing by the same autoimmune response. Recently, monoclonal antibodies which recognize human monocytes, T lymphocytes and T-cell subsets have become available. We have tested cultured ovine oligodendrocytes with a battery of monoclonal antibodies directed against human lymphocyte and monocyte antigens, seeking evidence for shared determinants, and present here data suggesting that shared determinants exist.

All antibody clones were produced by hybridization of spleen cells from mice sensitized to human T lymphocytes with P3 mouse myeloma. Antibodies were generated by intraperitoneal passage of cloned hybrid cells in mice⁶. Oligodendrocytes were isolated from the white matter of 4–6-month-old lambs as described previously^{7,8}. Trypsin was used in the isolation procedure. Cells were plated on glass chamber slides and maintained in culture for up to 40 days (for details see Fig. 1 legend); medium was changed twice weekly. Cultured cells were identified as oligodendrocytes by ultrastructural, biochemical and immunological criteria⁹. Cultured oligodendrocytes were washed free of medium and pre-reacted with normal sheep serum (NSS) as a source of excess sheep IgG. All monoclonal antibodies were screened at a concentration of 5 μg IgG per ml, the cultures were washed, and binding of the monoclonal antibodies was assessed by adding goat anti-mouse IgG labelled with fluorescein isothiocyanate (FITC).

As shown in Fig. 1, cultured oligodendrocytes pretreated with 1:500 NSS, exposed to OKT8 and then to 1:10 goat anti-mouse IgG–FITC, gave distinct membrane-bound fluorescence with scattered irregular dots outlining the cell cytoplasm and processes. This pattern of fluorescence resembles that seen when human suppressor lymphocytes are reacted with OKT8. A similar fluorescence pattern was obtained with OKT3 and OKM1. In contrast, OKT4, OKT5, OKT6, OKT10 or OKI1 showed no fluorescence. Control cultures incubated with goat anti-mouse IgG–FITC alone were negative, as was ascitic fluid (produced by injecting P3 mouse myeloma cells into mice), even at 2 mg ml⁻¹ IgG. In dilution experiments, using the same 1:10 dilution of goat anti-mouse FITC, OKT3 became negative at concentrations below 2.5 μg ml⁻¹ IgG, but OKT8 and OKM1 remained positive at 1.25 μg ml⁻¹ IgG. Table 1 gives the IgG subclasses of the monoclonal antibodies used, their specificities for human cell types, and summarizes the data on membrane-bound fluorescence.

As immunoglobulins can bind nonspecifically to Fc receptors on cells or specifically through their Fab portion, it was possible that monoclonal antibody binding was occurring through the Fc

Table 1 Binding of monoclonal antibodies to ovine oligodendrocytes

Monoclonal antibody	IgG class	Human specificity	Ovine oligodendrocyte binding at $5 \mu\text{g ml}^{-1}$ in the presence of	
			1:500 NSS	1:10 NSS
OKT3*	IgG2a	T/effector	+	—
OKT4†	IgG2b	T/helper	—	—
OKT5‡	IgG1	T/suppressor	—	ND
OKT6†	IgG1	Thymocytes	—	—
OKT8†	IgG2a	T/suppressor	+++	++
OKM1*‡	IgG2b	Monocytes	++	—
OKI1‡	IgG2a	Anti-Ia	—	—
OKT10‡	IgG1	Anti-thymocyte	—	—

ND, not determined.

* Commercially available from Ortho Laboratories at $25 \mu\text{g ml}^{-1}$ of mouse IgG. † Commercially available from Ortho Laboratories at $50 \mu\text{g ml}^{-1}$ of mouse IgG. ‡ Ascitic fluid gift of Dr P. C. Kung obtained at 2 mg ml^{-1} of mouse IgG.

region, particularly as oligodendrocytes have been shown to bear Fc receptors^{10,11}. Sheep red blood cells (SRBC) coated with mouse anti-SRBC were shown to form rosettes with our oligodendrocyte population. This Fc rosetting was totally inhibited by preincubating oligodendrocytes with a 1:10 dilution of NSS. Human lymphocytes to which the mouse monoclonal antibody OKT8 had been bound specifically also formed rosettes with oligodendrocytes (Fig. 2). This Fc receptor binding was also blocked by preincubating the oligodendrocytes with a 1:10 dilution of NSS (Fig. 2). Direct binding of OKT3, OKT8 and OKM1 to oligodendrocytes was therefore re-evaluated following a preincubation of the oligodendrocytes in 1:10 NSS. OKT3 and OKM1 binding was no longer seen, indicating that their binding might have occurred through Fc. Although Fc receptor binding is known to depend on IgG subclass when IgG and Fc receptors are from the same species¹², we found no correlation

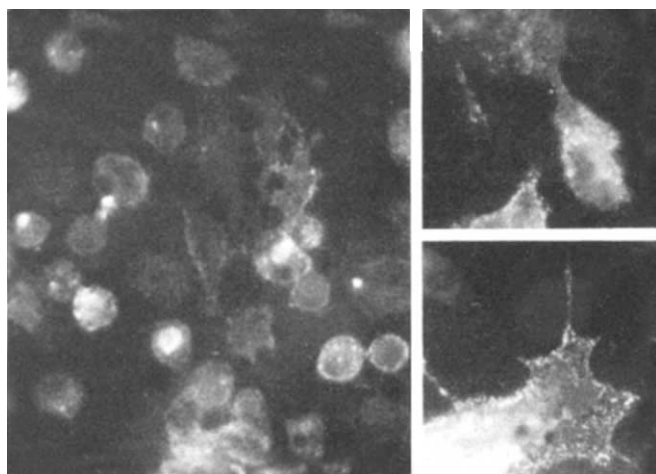


Fig. 1 Oligodendrocytes isolated from lamb brain (as described in ref. 8) were cultured (in multiple-well glass slides) in Dulbecco's modified minimum essential medium (with 20% horse serum added) at 37°C in humidified air containing 5% CO_2 . After 2–3 weeks in culture, the slides were washed three times with Ca^{2+} – Mg^{2+} –free Hanks' balanced salt solution containing 1% bovine serum albumin, pH 7.4 (HBSS + 1% BSA). They were incubated for 15 min with a 1:500 dilution of heat-inactivated normal sheep serum in HBSS + 1% BSA. The preincubation solution was then aspirated off and mixed with the monoclonal antibody OKT8 (Ortho) up to a final concentration of $5 \mu\text{g ml}^{-1}$ of mouse IgG. The mixture was added back to the cells and allowed to react for 30 min at 0°C . After three washes with HBSS + 1% BSA, a 1:10 dilution of goat anti-mouse IgG–FITC was added for 30 min at 0°C . The goat anti-mouse IgG–FITC (Cappel) was diluted with one part heat-inactivated normal goat serum and eight parts HBSS + 1% BSA. Slides were then washed four times, fixed with 4% paraformaldehyde in HBSS at 0°C , mounted and observed under a Leitz Ortholux II UV microscope. Pictures were taken with an Orthomat camera.

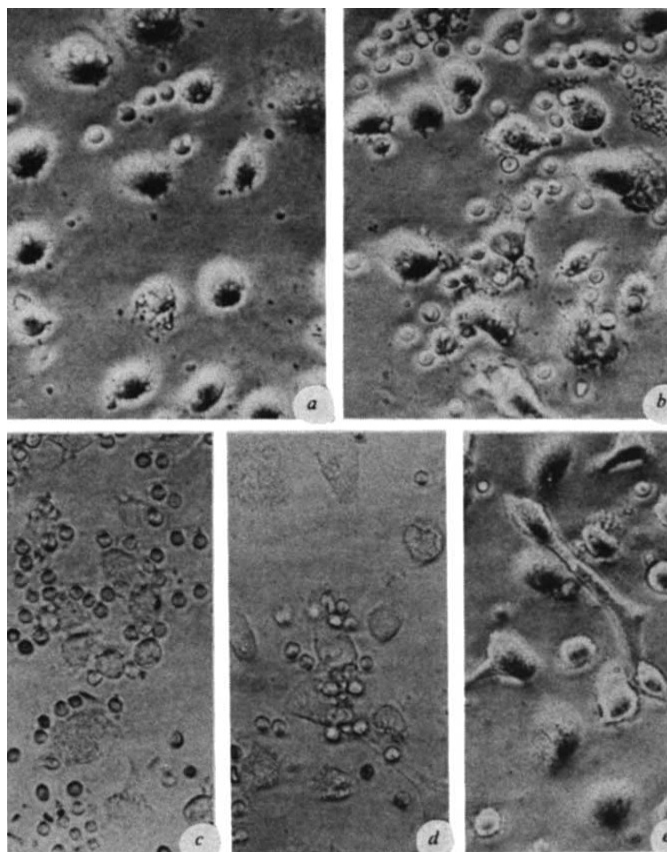


Fig. 2 Blocking effect of sheep serum on Fc binding of the monoclonal antibody OKT8. Human mononuclear cells were separated from peripheral blood by Ficoll–Hypaque flotation¹³. T lymphocytes were then isolated by the nylon wool technique¹⁴. After washing, T cells were incubated at 4°C for 30 min with a $5 \mu\text{g ml}^{-1}$ solution of OKT8 in HBSS + 1% BSA and washed three times in the cold. 20% of the T lymphocytes were shown by immunofluorescence to have bound OKT8. T cells coated with OKT8 were then added to oligodendrocytes in culture. After centrifugation at 200g for 5 min, incubation for 30 min and gentle washing, cultures were observed for the presence of T cells bound to oligodendrocytes. In *a*, uncoated lymphocytes have been added and no binding is seen. In *b*, T cells precoated with OKT8 have been added to washed oligodendrocytes; the presence of free Fc receptor is shown by binding of the T cells to the oligodendrocytes. *c*, *d* Demonstrate reduced binding after preincubating oligodendrocytes with 1:500 (*c*) and 1:50 (*d*) NSS. *e* Shows that a 1:10 dilution of NSS in the preincubation step totally blocks binding to Fc receptors.

between mouse IgG subclass and monoclonal antibody binding to ovine oligodendrocyte Fc receptor. This might be attributable to species differences, to the presence of aggregates in some of the OKT or to variability in the structure of the Fc regions of the monoclonal antibodies.

After preincubation with 1:10 NSS, OKT8 binding was still evident by immunofluorescence (Fig. 3), again in a pattern reminiscent of that seen when human suppressor T cells are reacted with OKT8. Ultracentrifugation of OKT8 at 100,000g for 1 h, to remove aggregated IgG, did not affect its binding to oligodendrocytes.

Galactocerebroside is an accepted marker for oligodendrocytes. Double labelling experiments were done using, in a first step, a mixture of OKT8 and rabbit anti-galactocerebroside antiserum and in the second step, goat anti-mouse IgG (FITC labelled) mixed with goat anti-rabbit IgG (rhodamine labelled). Using three-week-old cultures, 75% of cells labelled for galactocerebroside, 86% for the T8 antigen and 68% showed double labelling.

Freshly isolated ovine oligodendrocytes studied in suspension did not exhibit fluorescence with any of the antibodies tested.

Thus, the OKT8 reactive determinant, as well as the Fc binding site, seems to be lost during the isolation procedure and regenerated in culture. Failure of freshly prepared cells to bind could be a consequence of the trypsinization step used during isolation. Indeed, when 0.3 mg ml^{-1} of trypsin was added for 45 min at 37°C to human mononuclear cells in suspension or to ovine oligodendrocytes which had been maintained in culture, both lost their capacity to bind OKT8, OKT3 and OKM1. Serial experiments performed using OKT8 showed that antibody binding was first evident after 3 days in culture and was invariably present thereafter up to 40 days in culture, the last time point tested. Per cent positive cells varied between cultures but binding was always present in more than 50% of the cells.

We attempted to stain non-trypsinized sheep hepatocytes in suspension and cultured sheep fibroblasts using the same antibodies; neither cell type was positive with any monoclonal antibody. These results indicate that cross-reactivity of the monoclonal antibodies is at least relatively restricted to oligodendrocytes and does not reflect reactivity with globally distributed ovine antigens. Interestingly, sheep splenocytes in

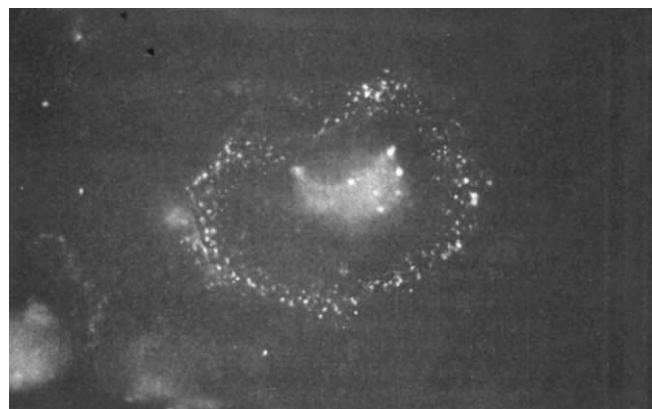


Fig. 3 Immunofluorescence pattern of oligodendrocytes stained with OKT8 in conditions described in Fig. 1 legend, except that a 1:10 dilution of normal sheep serum (shown to block Fc receptors) was used in the preincubation step.

suspension did not stain with any of the monoclonal antibodies. It would be of paramount interest to develop methods for maintaining human oligodendrocytes in culture and to test them for reactivity against OKT3, OKM1 and OKT8.

If determinants are shared by suppressor lymphocytes and oligodendrocytes, as the present data suggest, this could explain why both suppressor cells and oligodendrocytes disappear in multiple sclerosis. The nature of the putative multiple sclerosis-relevant antigen in brain may even be approachable through study of circulating lymphocytes.

This work was supported by National Multiple Sclerosis Society grants MS:RG 1223-B-3 and MS:RG 1130-C-16. Monoclonal antibodies (in mouse ascites fluid) were purchased from Ortho Laboratories (OKT3, OKT4, OKT6 and OKT8) or were provided by Dr P. C. Kung (OKM1, OKT5, OKT10 and OKI1). Anti-galacto cerebroside antiserum was a gift from Dr R. Lisak.

Received 28 August; accepted 20 October 1981.

- Charcot, J. M. *Gaz. Hôp. Paris* **41**, 554–566 (1868).
- McAlpine, D., Compston, N. D. & Lumsden, C. E. (eds) *Multiple Sclerosis* (Livingston, Edinburgh, 1955).
- Antel, J. P., Arnason, B. G. W. & Medof, M. E. *Ann. Neurol.* **5**, 338–342 (1979).
- Reinherz, E. L. *et al. New Engl. J. Med.* **303**, 125–129 (1980).
- Bach, M. A. *et al. Lancet* **ii**, 1221–1222 (1980).
- Reinherz, E. L., Kung, P. C., Goldstein, G. & Schollossman, S. F. *J. Immun.* **124**, 3, 1301–1307 (1980).
- Szuchet, S., Arnason, B. G. W. & Polak, P. J. *J. Neurosci. Meth.* **3**, 7–19 (1980).
- Szuchet, S. & Stefansson, K. *Adv. cell. Neurobiol.* **1**, 313–346 (1980).
- Szuchet, S., Stefansson, K., Wollmann, R. L., Dawson, G. & Arnason, B. G. W. *Brain Res.* **200**, 151–164 (1980).
- Goust, J. M., Banik, N., Bethea, A. & Sullivan, L. M. *Trans. Am. Neurol. A* (in the press).
- Ma, B. I., Joseph, B., Walsh, M., Potuim, P. F. & Tourtellotte, W. W. *Ann. Neurol.* **9**, 371–377 (1981).
- Diamond, B. & Scharff, M. D. *J. Immun.* **125**, 631–634 (1980).
- Boyum, A. *Scand. J. clin. Lab. Invest.* **21**, Suppl. 97, 77–89 (1968).
- Greenwalt, T. S., Gajewski, M. & McKenna, J. L. *Transfusion* **2**, 221–229 (1962).

T-cell mitogens cause early changes in cytoplasmic free Ca^{2+} and membrane potential in lymphocytes

R. Y. Tsien*, T. Pozzan† & T. J. Rink

Physiological Laboratory, University of Cambridge, Cambridge CB2 3EG, UK

† Istituto di Patologia Generale, Via Loredan 16, 35100 Padova, Italy

One of the best models for studying the control of mammalian cell growth and proliferation is the response of lymphocytes to mitogenic agents^{1–4}, which stimulate growth, DNA synthesis and division. How mitogens work remains obscure, although it has been hypothesized that they increase cytoplasmic free calcium, $[\text{Ca}^{2+}]_i$, as a trigger for the cascade of intracellular processes necessary for proliferation^{3–6}. However, lymphocyte $[\text{Ca}^{2+}]_i$ has not previously been measurable. A new technique⁷ for loading a novel Ca^{2+} -specific indicator⁸ into the cytoplasm of intact small cells has now made possible the first direct measurements of $[\text{Ca}^{2+}]_i$ in mouse thymocytes and pig node lymphocytes. We show here that lectins known to stimulate T cells raise average $[\text{Ca}^{2+}]_i$ approximately twofold within a few minutes. Deprivation of external Ca^{2+} or elevation of cyclic AMP, conditions known to inhibit mitogenesis, prevented the $[\text{Ca}^{2+}]_i$ response. Rises in $[\text{Ca}^{2+}]_i$ were accompanied by hyperpolarization of the membrane potential, apparently due to a Ca^{2+} -activated K^+ conductance. The co-carcinogen 12-*O*-tetradecanoylphorbol-13-acetate (TPA) seems to stimulate cell functions normally activated by Ca^{2+} .

$[\text{Ca}^{2+}]_i$ was measured with a fluorescent indicator (structure 3b of ref. 8, now called 'quin2') which shows a fivefold fluorescence enhancement on binding Ca^{2+} . Quin2 has a 1:1 stoichiometry of Ca^{2+} binding, with an apparent dissociation constant⁹ of 115 nM in the presence of 130 mM K^+ , 20 mM Na^+ , 1 mM free Mg^{2+} , at pH 7.05 and 37°C , conditions chosen to mimic lymphocyte cytoplasm (refs 10, 11 and T.P., T.J.R. and R.Y.T., in preparation). The indicator was trapped inside the cells by incubating them with its membrane-permeant intracellularly hydrolysable acetoxymethyl ester⁷. Available tests suggest that the quin2 is free in the cytoplasm, not significantly contained in organelles such as mitochondria, and neither excluded from nor accumulated in the nucleus, so that its signals can be taken to measure cytoplasmic free calcium⁹. The signals are calibrated by releasing the indicator from the cells with digitonin or Triton X-100 at the end of the experiment and recording the fluorescence at 0% and 100% Ca saturation, that is, $<1 \text{ nM}$ and 1 mM $[\text{Ca}^{2+}]$. From these fluorescence levels, $F_{\min}$ and $F_{\max}$, the $[\text{Ca}^{2+}]_i$ corresponding to fluorescences F from trapped intracellular dye, are calculated by the equation

$$[\text{Ca}^{2+}] = (115 \text{ nM})(F - F_{\min}) / (F_{\max} - F)$$

Parallel measurements of lymphocyte membrane potential (V_m) were made with a potential-sensitive oxonol dye as previously described¹².

Figure 1 shows fluorescence signals from quin2-loaded lymphocytes. The indicated resting $[\text{Ca}^{2+}]_i$ had a mean $\pm$ s.e. of $123 \pm 6 \text{ nM}$ ($n = 27$). These levels are consistent with those measured in other cell types¹³ and argue against any gross toxicity. Further evidence against toxicity was that (1) loaded cells maintained high ($>90\%$) eosin exclusion; (2) ATP levels⁹ remained within 30% of normal with up to 3 mM intracellular quin2; (3) capping of surface ligands was unaffected by quin2 (ref. 14); (4) membrane potential responses of quin2-loaded mouse thymocytes were similar to those of normal cells

* Present address: Department of Physiology-Anatomy, University of California, Berkeley, California 94720, USA.

(compare Figs 2 and 3b); and (5) thymocytes and spleen cells loaded with quin2 and then treated with concanavalin A (Con A) at normal optimal doses can undergo mitogenic transformation as assessed by thymidine uptake (T. R. Hesketh, personal communication). All our results pertain to cells washed free from ester hydrolysis products after the loading period; leaving the hydrolysis by-products in the suspension may have long-term deleterious effects.

Although quin2 is a Ca^{2+} chelator, increasing its loading does not affect steady-state $[\text{Ca}^{2+}]_i$. Atomic absorption measurement of total Ca shows that during loading, the cells take up sufficient Ca from the medium approximately to half-saturate the trapped quin2 and keep $[\text{Ca}^{2+}]_i$ near 120 nM. Loading in Ca^{2+} -free medium does reduce $[\text{Ca}^{2+}]_i$ to as low as 16 nM (ref. 9).

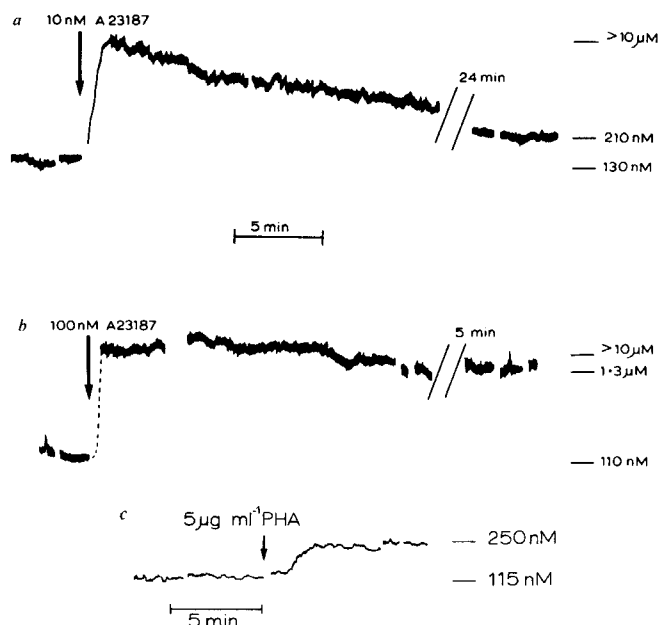


Fig. 1 Resting $[\text{Ca}^{2+}]_i$ levels and responses to A23187 and PHA. *a, b*, Lymphocytes from the mesenteric nodes of pigs were prepared as previously described³⁶. Quin2 acetoxymethyl ester (50 μM) was added to a cell suspension containing about 10^8 cells per ml and incubated for 20 min. The suspension was then diluted 10-fold and incubated for a further 40–60 min. The resulting quin2 loading was typically 2 mmol per litre of cells; higher or lower loadings could be achieved by altering the ester concentration. After loading, the cells were centrifuged at 1,000g for 3–4 min and resuspended in fresh RPMI-1640 at $\sim 10^7$ per ml and kept at room temperature. Cells usually lost <5% of their dye per hour. For measurement of fluorescence, 1–2 ml of this stock suspension was centrifuged for a few seconds at 14,000g and the cells resuspended in 2 ml of a physiological saline and transferred to the cuvette, the whole resuspension procedure taking about 1 min. Unless otherwise specified, the physiological saline consisted of (mM): NaCl, 145; KCl, 5; Na_2HPO_4 , 1; CaCl_2 , 1; MgSO_4 , 0.5; glucose, 5; Na HEPES, 10; pH 7.40 at 37 °C. This simplified saline was routinely used instead of RPMI-1640 because the latter contains many absorbing and fluorescing substances; however, qualitatively similar $[\text{Ca}^{2+}]_i$ signals (see *c*) could be obtained in the full culture medium. Typical measurement conditions, with 2 mM quin2 content in $0.5-1 \times 10^6$ cells per ml, are equivalent to about 1–2 μmol dye per litre of suspension. Quin2 fluorescences were recorded with Perkin-Elmer MPF-44 spectrofluorimeters in cuvettes thermostatted to 37 °C. Excitation and emission wavelengths were 339 and 492 nm with 4 and 10 nm bandwidths. 10 or 100 nM A23187 added from dimethyl sulphoxide (DMSO) stock solutions as indicated. Final DMSO concentrations were 0.1% (v/v); control experiments showed no effect of DMSO up to 1%. *c*, PHA (Sigma) was added as indicated to mouse thymocytes teased from thymus glands of 5–8-week-old BALB/c mice and loaded with 2 mM quin2 content as in *a, b*. This experiment was done in RPMI-1640 medium containing 1 mM free Ca^{2+} and 0.2 mM Ca^{2+} -EGTA. Although PHA is less effective at stimulating thymidine uptake in thymocytes than in mature T cells, it does induce the early metabolic changes characteristic of mitogenesis^{37,38}. In this and Figs 2 and 3, fluorescence was calibrated in terms of $[\text{Ca}^{2+}]_i$ as described in the text from F_{max} and F_{min} determined after quin2 release by 50 μM digitonin or 0.02% (v/v) Triton X-100. Cells not loaded with quin2 gave no fluorescence responses to ionophores or lectins. Gaps in the trace denote additions to or stirring of the suspension. The synthesis of quin2 is described in ref. 8. Quin2/AM was prepared from quin2 by the general procedure of ref. 7. Some experiments used batches of quin2/AM chromatographically purified and crystallized by Dr G. A. Smith, but with no difference in results from less exhaustively purified material.

Two different mitogenic calcium ionophores, A23187 (ref. 15) and ionomycin¹⁶, each proved very potent in nanomolar concentrations at raising $[\text{Ca}^{2+}]_i$. Fig. 1a shows that 10 nM A23187, a submitogenic dose, increased $[\text{Ca}^{2+}]_i$ to a peak of $\geq 1 \mu\text{M}$, from which it gradually decayed back nearly to resting level during 1 h. The decay may be due to slow disappearance of ionophore from solution¹⁷, and could explain why 10 nM is a submitogenic dose³. A second addition of 10 nM A23187 can produce a renewed $[\text{Ca}^{2+}]_i$ rise. A single application of 100 nM A23187, the optimal mitogenic concentration³, is sufficient to keep $[\text{Ca}^{2+}]_i$ above 1 μM for at least 40 min (Fig. 1b).

T lymphocytes can also be stimulated by the lectins phytohaemagglutinin (PHA) and Con A. At typically mitogenic doses both lectins give similar quin2 responses from thymocytes in 1 mM extracellular Ca^{2+} . A typical PHA response is shown in Fig. 1c. After a delay of 30–60 s, the fluorescence rises over 2–3 min by an amount corresponding to a 1.5- to 2.5-fold increase in average $[\text{Ca}^{2+}]_i$. Increasing the dye loading has no effect on the final apparent level of $[\text{Ca}^{2+}]_i$, which is 200–300 nM, but slows the response slightly. The elevation of $[\text{Ca}^{2+}]_i$ is maintained for at least tens of minutes with optimally mitogenic doses of Con A. In nominally Ca^{2+} -free media, Con A gives no increase in $[\text{Ca}^{2+}]_i$, which fits with the dependence^{3,15,18,19} of lectin mitogenicity on extracellular Ca^{2+} . If Ca^{2+} is then added back to the Con A medium, $[\text{Ca}^{2+}]_i$ rises immediately. Agents which raise intracellular cyclic AMP levels are known to prevent or abort mitogenic stimulation^{5,20,21}; the $[\text{Ca}^{2+}]_i$ response in thymocytes to Con A is antagonized by such agents, for example 100 μM theophylline, 10 mM Ro20/1724 (another phosphodiesterase inhibitor²¹), 0.25 μM dibutyryl cyclic AMP, or preincubation with 1 $\mu\text{g ml}^{-1}$ cholera toxin. By contrast, no effects on $[\text{Ca}^{2+}]_i$ of millimolar exogenous cyclic GMP or its dibutyryl or 8-bromo derivatives have yet been detected.

These lectin-induced rises in average $[\text{Ca}^{2+}]_i$ may seem small. However, any influence of $[\text{Ca}^{2+}]_i$ on cell growth is probably exerted over several hours³, giving time for a small activation to have a large cumulative effect. Moreover, not all the cells in these unfractionated populations are identically responsive to lectin^{3,22–24}. If, say, half the cells remain at resting $[\text{Ca}^{2+}]_i$, at which level the quin2 is about 50% saturated with Ca^{2+} , the quin2 in the responsive cells would have to reach 80–90%

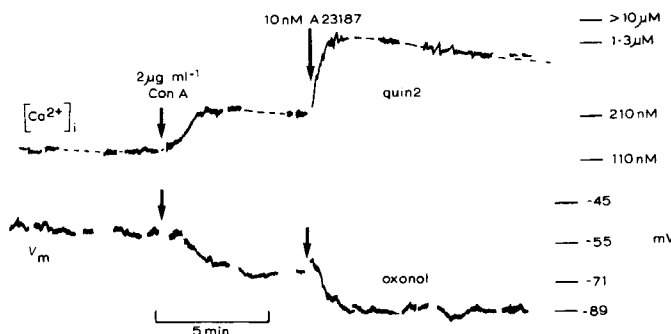


Fig. 2 Effect of Con A (Miles-Yeda, crystallized three times) on $[\text{Ca}^{2+}]_i$ and membrane potential, V_m . Mouse thymocytes containing 3 mM quin2 were in the standard physiological saline, to which was added 0.1 μM oxonol. Upper trace shows the quin2 signal, lower trace the oxonol fluorescence (excitation 540 nm, emission 580 nm) from another aliquot of the same batch of quin2-loaded cells. During each recording, the wavelengths were intermittently changed to verify that the other dye was responding normally, hence the large number of gaps in each trace. The V_m calibration was obtained after application of A23187 by adding KCl to raise external K^+ in steps and assuming that the increase in K^+ permselectivity was by then sufficient to bring V_m close to the K^+ equilibrium potential, E_K . The calibration marks relate oxonol fluorescence to E_K for each external K^+ concentration. The value of resting potential, -50 mV , estimated from oxonol fluorescence, is in reasonable agreement with independent assessments using the distribution of triphenylmethylphosphonium cation (TPMP⁺) in other lymphocyte types^{39,40}. However, the lectin-induced hyperpolarizations have not been reported previously. TPMP⁺ equilibration takes 0.5–1 h and is thus not suitable for following early changes. Cyanine dyes¹² are unsuitable probes because they poison mitochondria, deplete lymphocyte ATP⁴¹ and raise $[\text{Ca}^{2+}]_i$ to a point where lectins produce no further increase.

saturation ($[Ca^{2+}]_i = 450\text{--}1,000\text{ nM}$) to account for the observed average rise to 65–70% saturation. The quin2 response does not come primarily from the $[Ca^{2+}]_i$ of macrophages, since the response was not reduced by depletion of macrophages by adherence to glass, nor was any fluorescence response to Con A observed from macrophages obtained by peritoneal lavage and loaded with quin2.

We have found that lectin-induced $[Ca^{2+}]_i$ rises are sufficient to cause at least one immediate cell response: a negative shift of membrane potential. Figure 2 compares the $[Ca^{2+}]_i$ rise and the hyperpolarization in a single batch of quin2-loaded cells responding to mitogenic doses of Con A. The membrane potential as assessed by oxonol fluorescence changes from its resting level of approximately -50 mV to -70 mV with an initial delay and S-shaped time course similar to the $[Ca^{2+}]_i$ rise. A23187 produces an immediate further increase in $[Ca^{2+}]_i$ and hyperpolarization. The hyperpolarization is likely to be due to a K^+ conductance because it is associated with increased responsiveness of V_m to external $[K^+]$; also, of the main intracellular ions, only K^+ has an equilibrium potential^{10,11,25} more negative than the resting potential. The hyperpolarization is also blocked by 0.5 mM quinine, an agent known to block Ca^{2+} -activated K^+ conductances¹⁷. Stimulation of electrogenic Na^+/K^+ pumping is probably not responsible because the hyperpolarization persists in the presence of 1 mM ouabain with only 1 mM external K^+ . The hyperpolarization is a consequence of the $[Ca^{2+}]_i$ rise, not vice versa, as in Ca^{2+} -free solution Con A induces neither a rise in $[Ca^{2+}]_i$ nor hyperpolarization, but when external Ca^{2+} is restored, the $[Ca^{2+}]_i$ rise and V_m hyperpolarization develop in parallel. Conversely, imposed hyperpolarization or depolarization²⁶ do not affect the Con A-induced $[Ca^{2+}]_i$ rise. The similarity of the V_m responses between cells with and without quin2 (Figs 2 and 3b) argues that the indicator does not greatly perturb the $[Ca^{2+}]_i$ response to Con A.

A counterexample to the correlations between known mitogenic potency and the ability to raise $[Ca^{2+}]_i$ was found in TPA, a weak mitogen by itself and a potent synergist with ionophores and lectins^{16,27–30}. Ten nanomolar TPA gradually lowers resting $[Ca^{2+}]_i$, suppresses Con A-induced rises in $[Ca^{2+}]_i$ (Fig. 3a) and can even partly antagonize an ionophore response (Fig. 3c). These effects may be due to stimulated Ca^{2+} pumping. Yet TPA hyperpolarizes the cells by itself and enhances a Con A-induced hyperpolarization (Fig. 3b). The transience of the TPA hyperpolarization in Fig. 3b can be explained by the eventual fall in $[Ca^{2+}]_i$ counteracting the initial direct stimulation or increase in Ca^{2+} sensitivity of the K^+ conductance. All these actions of TPA are detectable in tens of seconds or minutes and are near-maximal at 10 nM , a concentration comparable with the reported dissociation constants for lymphocyte TPA-binding sites³¹. We suggest that TPA fairly directly stimulates a variety of normally $[Ca^{2+}]_i$ -dependent processes, such as the Ca pump, K^+ conductance and events leading to mitogenesis. Analogous stimulations might explain many actions of TPA on other cell types^{32–34}, including the aggregation of platelets while lowering $[Ca^{2+}]_i$ (S. W. Smith, T.J.R. & R.Y.T., unpublished data). Interestingly, the high-affinity binding of phorbol diesters to cellular receptors is enhanced by $[Ca^{2+}]_i$ in the submicromolar range³⁵, suggesting that TPA binding may reciprocally increase the Ca^{2+} affinities of those receptors.

These results support the view that $[Ca^{2+}]_i$ is a primary messenger in lymphocyte mitogenesis, although we cannot say whether it is the only initial messenger without further studies such as analysis of lymphocyte heterogeneity and correlation of long-term $[Ca^{2+}]_i$ with mitogenesis. The surprising ability of the co-mitogen TPA to lower $[Ca^{2+}]_i$ can be rationalized by its stimulation of normally Ca^{2+} -dependent processes. Changes in $[Ca^{2+}]_i$ sensitivity of cellular processes, as well as modest sustained rises in $[Ca^{2+}]_i$, may be of more general significance in the control of normal and malignant cell proliferation.

This work was supported by SRC and CNR grants. T.P. held an EMBO Long-term Fellowship. We thank Dr G. A. Smith for samples of specially purified quin2 acetoxymethyl ester ('quin2/AM'), and Dr T. R. Hesketh for permission to cite unpublished results.

Received 15 July; accepted 24 November 1981.

- Oppenheim, J. J. & Rosenstreich, D. L. *Prog. Allergy* **20**, 65–194 (1976).
- Wedner, H. J. & Parker, C. W. *Prog. Allergy* **20**, 195–300 (1976).
- Metcalfe, J. C., Pozzan, T., Smith, G. A. & Hesketh, T. R. *Biochem. Soc. Symp.* **45**, 1–26 (1980).
- Hume, D. A. & Weidemann, M. J. *Mitogenic Lymphocyte Transformation* (Elsevier, Amsterdam, 1980).
- Coffey, R. G., Hadden, E. M. & Hadden, J. W. *J. Immun.* **119**, 1387–1394 (1977).
- Whitfield, J. F. *et al. Ann. N.Y. Acad. Sci.* **339**, 216–240 (1980).
- Tsien, R. Y. *Nature* **290**, 527–528 (1981).
- Tsien, R. Y. *Biochemistry* **19**, 2396–2404 (1980).
- Pozzan, T., Rink, T. J. & Tsien, R. Y. *J. Physiol., Lond.* **318**, 12P–13P (1981); *J. Cell Biol.* (submitted).
- Lichtman, M. A., Jackson, A. H. & Peck, W. A. *J. cell. Physiol.* **80**, 383–396 (1972).
- Pozzan, T., Rink, T. J. & Tsien, R. Y. *J. Physiol., Lond.* **319**, 102P (1981).
- Rink, T. J., Montecucco, C., Hesketh, T. R. & Tsien, R. Y. *Biochim. biophys. Acta* **595**, 15–30 (1980).
- Alvarez-Leefmans, F. J., Rink, T. J. & Tsien, R. Y. *J. Physiol., Lond.* **315**, 531–548 (1981).
- Pozzan, T., Arslan, P., Tsien, R. Y. & Rink, T. J. *J. Cell Biol.* (submitted).
- Maino, V. C., Green, N. M. & Crumpton, M. J. *Nature* **251**, 324–327 (1974).
- Lyall, R. M., DuBois, J. H. & Crumpton, M. J. *Biochem. Soc. Trans.* **8**, 720–721 (1980).
- Lew, V. L. & Ferreira, H. G. *Curr. Topics Membranes Transport* **10**, 217–277 (1978).
- Freedman, M. H. in *Cell Biology and Immunology of Leukocyte Function* (ed. Quastel, M.R.) (Academic, New York, 1979).
- Bard, E., Colwill, R., L'Anglais, R. & Kaplan, J. G. *Can. J. Biochem.* **56**, 900–904 (1978).
- Watson, J. J. *Immun.* **117**, 1656–1663 (1976).
- DeRubertis, F. R. & Zenser, T. *Biochim. biophys. Acta* **428**, 91–103 (1976).
- Whitesell, R. R., Hoffman, L. H. & Regen, D. M. *J. biol. Chem.* **252**, 3533–3537 (1977).
- Gerhart, S., Mills, G., Monticone, V. & Paetkau, V. *J. Immun.* **117**, 1314–1319 (1976).
- Wagner, H. *et al. Immun. Rev.* **51**, 215–255 (1980).
- Iversen, J. G. *J. cell. Physiol.* **89**, 267–276 (1976).
- Montecucco, C., Rink, T. J., Pozzan, T. & Metcalfe, J. C. *Biochim. biophys. Acta* **595**, 65–70 (1980).
- Whitfield, J. F., MacManus, J. P. & Gillan, D. J. *J. cell. Physiol.* **82**, 151–156 (1973).
- Mastro, A. M. & Mueller, G. C. *Exptl Cell Res.* **88**, 40–46 (1974).
- Wang, J. L., McClain, D. A. & Edelman, G. M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1917–1921 (1975).
- Rosenstreich, D. L. & Mizel, S. B. *J. Immun.* **123**, 1749–1754 (1979).
- Estensen, R. D., DeHoogh, D. K. & Cole, C. F. *Cancer Res.* **40**, 1119–1124 (1980).
- Schleimer, R. P., Gillespie, E. & Lichtenstein, L. M. *J. Immun.* **126**, 570–574 (1981).
- Estensen, R. D. *et al. in Control of Proliferation in Animal Cells* (eds Clarkson, B. & Baserga, R.) (Cold Spring Harbor Press, New York, 1974).
- Malaise, W. J. *et al. Cancer Res.* **40**, 3827–3831 (1980).
- Dunphy, W. G., Kochenburger, R. J., Castagna, M. & Blumberg, P. M. *Cancer Res.* **41**, 2640–2647 (1981).

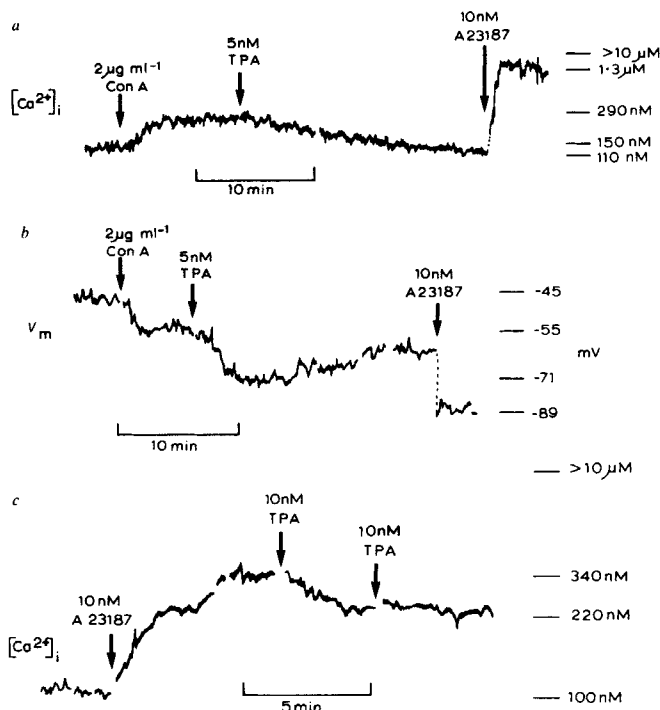


Fig. 3 Effects of TPA (Sigma). *a*, Mouse thymocytes containing 2 mM quin2. After elevation of $[Ca^{2+}]_i$ by Con A, $[Ca^{2+}]_i$ is reduced by 5 nM TPA to below the resting level. Finally 10 nM A23187 raises $[Ca^{2+}]_i$ to above $1\text{ }\mu\text{M}$. *b*, Mouse thymocytes from the same preparation as *a* but not loaded with quin2. Signal from $0.1\text{ }\mu\text{M}$ oxonol. After a Con A hyperpolarization, 5 nM TPA causes a further, incompletely maintained hyperpolarization. V_m was calibrated as for Fig. 2. *c*, Pig lymphocytes loaded with 2.5 mM quin2. 10 nM TPA partially reversed the rise in $[Ca^{2+}]_i$ caused by A23187. A second addition of 10 nM TPA had no further effect. The smallness of the initial response to 10 nM A23187 may have been due to reduced potency of the particular stock solution used.

36. Hesketh, T. R., Smith, G. A., Houslay, M. D., Warren, G. B. & Metcalfe, J. C. *Nature* **267**, 490-494 (1977).
37. Resch, K., Ferber, E., Prester, M. & Gelfand, E. W. *Eur. J. Immun.* **6**, 168-173 (1976).
38. Gronik, K.-O. & Andersson, J. *Immun. Rev.* **51**, 35-59 (1980).
39. Kiefer, H., Blume, A. J. & Kaback, H. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2200-2204 (1980).
40. Deutsch, C. J., Holian, A., Holian, S. K., Daniele, R. P. & Wilson, D. F. *J. cell. Physiol.* **99**, 79-94 (1979).
41. Montecucco, C., Pozzan, T. & Rink, T. J. *Biochim. biophys. Acta* **552**, 552-557 (1979).

Lower fidelity of RecA protein catalysed homologous pairing with a superhelical substrate

Chanchal DasGupta & Charles M. Radding

Departments of Human Genetics and Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, Connecticut 06510, USA

The heteroduplex joint, a splice containing paired strands from each of two DNA molecules is a key feature of homologous genetic recombination^{1,2}. The formation of such a lap joint ensures that heterologous chromosomes do not recombine; presumably the degree of homology determines the frequency of crossing over between related but non-identical chromosomes. On the other hand, according to the current ideas on recombination, the formation of heteroduplex joints is not a stringent process. Genetic evidence supports the view that the classical phenomena of meiotic gene conversion and aberrant meiotic segregation result in part from the inclusion of mismatched base pairs in heteroduplex joints, and the subsequent correction of some of these mismatched pairs before replication³⁻⁵. Thus meiotic gene conversion signals one kind of departure from perfectly faithful pairing. The association of increased mutation frequencies with crossing-over is another kind of departure from fidelity in recombination⁶⁻⁸. Little is known about the effect of imperfect homology either on the initial pairing of DNA molecules, or on the formation and extension of heteroduplex joints. The nucleotide sequences of phages Φ X174 and G4 are related, but differ by 33% of bases in the coding regions. Here we show using combinations of Φ X174 and G4 DNA, that *Escherichia coli* RecA protein catalyses the formation of joint molecules with many mismatched base pairs, but only if one of the molecules is superhelical. By contrast, in the absence of RecA protein, superhelicity does not cause Φ X174 and G4 DNA to form D-loops spontaneously at any temperature between 37 and 75 °C. These observations focus attention on the role of RecA protein in unwinding DNA, and on superhelicity as a factor that can lessen the fidelity of homologous pairing.

Recent studies of RecA protein, which is essential for genetic recombination of *E. coli*^{9,10} have shown that this single purified protein can promote two central steps of genetic recombination *in vitro*, namely synapsis, the pairing of homologous molecules, and strand exchange, which produces long heteroduplex joints¹¹⁻¹³. Synapsis is experimentally separable from the formation of heteroduplex joints: RecA protein can pair a single strand with homologous duplex DNA in the absence of a free end in any of the strands¹¹; and *E. coli* topoisomerase I, acting on these intermediates can interwind a single strand with its complement in duplex DNA¹². When a free end exists, for example in the combination of linear duplex DNA and circular single-stranded DNA, RecA protein causes the formation of long heteroduplex joints¹¹ by driving strand exchange in one direction¹⁴⁻¹⁶. RecA protein can also cause the uptake of a linear single strand by superhelical or non-superhelical DNA to form a particular kind of heteroduplex joint called a D-loop^{17,18}.

In vivo, both synapsis and heteroduplex formation presumably contribute to the fidelity of homologous pairing. In experiments on the reactions promoted *in vitro* by RecA protein, we have observed effects of imperfect homology on both synapsis and strand exchange^{12,15}. RecA protein makes joint

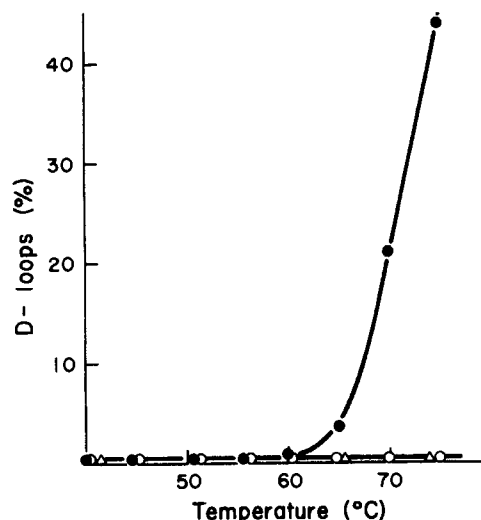


Fig. 1 Spontaneous formation of D-loops plotted against temperature. Superhelical DNA, and fragments of single-stranded DNA, each at 80 μ M, were heated for 40 min at the temperature indicated in 20 μ l reaction mixtures containing 10 mM Tris (pH 7.5), 1 mM EDTA and 200 mM NaCl. D-loops were assayed according to Beattie *et al.*²⁵. The superhelical DNA in all reaction mixtures was from Φ X174. ●, Φ X174 single-stranded fragments; ○, G4 single-stranded fragments; △, no single-stranded DNA.

Table 1 Superhelicity promotes the pairing of Φ X174 and G4 DNA by RecA protein

Double-stranded DNA	Single-stranded DNA		Joint molecules (%)
(a) Linear	Fragments	Circular	
G4	G4		42
	Φ X	G4	32
		Φ X	0.1
		Φ X	0.4
Φ X	Φ X		36
	G4	Φ X	12
		G4	0
		G4	0.2
(b) Superhelical	Fragments	Circular	
G4	G4		69
	Φ X	G4	26
		Φ X	21
		Φ X	0
Φ X	Φ X		70
	G4	Φ X	31
		G4	17
		G4	0
(c) Nicked circular	Fragments		
G4	G4		43
	Φ X		6
	Φ X		38
	G4		6

Duplex DNA, 4.2 μ M, and single-stranded DNA, 6.0 μ M, were incubated for 30 min at 37 °C in 20 μ l reaction mixtures containing 31 mM Tris (pH 7.5), 100 μ g/ml bovine serum albumin, 25 mM $MgCl_2$, 1.2 mM ATP and 3.0 μ M RecA protein. When the duplex DNA was superhelical, the reaction was stopped by adding 130 μ l of 25 mM EDTA (pH 8.5) and 0.5% Sarkosyl. Aliquots of 40 μ l were diluted in 2 ml of 1.5 M NaCl/0.15 M Na citrate, held at 41 °C for 4 min, diluted in cold NaCl/Na citrate, and filtered through nitrocellulose¹¹. When the duplex DNA was nicked circular or linear DNA, we stopped the reaction by adding 100 μ l of 25 mM EDTA (pH 8.5) and diluted it in cold 1.5 M NaCl/0.15 M Na citrate before filtration.

Table 2 Sequences of the 5' ends of the viral strand Φ X174 and G4 DNA each beginning at a different *Pst*I site

G4 DNA													
6			6/12					29					
G	GAA	TTG	CTG	TTG	CTA	ACT	TAC	GAA	TTA	AAT	CGA	AGT	
9/10													
GGA	CTG	CTG	GTG	GAA	AAT	GAG	GAA	ATT	CAA	TCT	CAA		
Φ X174 DNA													
15				15/23									
G	AGT	TTT	ATC	GCT	TCC	ATG	ACG	CAG	AAG	TTA	ACA	CTT	
16													
TCG	GAT	ATT	TCT	GAT	GAG	TCG	AAA	AAT	TAT	CTT	CAT		

Sequences are taken from Godson^{21,22}. For G4 DNA, the sequence shown is that of the viral strand, proceeding 5' to 3' from the *Pst*I site in gene A; for Φ X174 it is the sequence from a different *Pst*I site in gene A. Letters in bold type represent differences between the corresponding sequences of Φ X174 and G4. The brackets separate continuous runs of homologous bases from regions containing differences, and the numbers above the brackets indicate the number of perfectly matched bases.

molecules from mixed pairs of circular single-stranded DNA and linear duplex DNA of fd and M13, whose nucleotide sequences differ by 3% of bases^{19,20}. However, even these relatively sparse mismatches can stop the unidirectional growth of heteroduplex joints¹⁵. When we did the same experiment with pairs of Φ X174 and G4 DNA, no joint molecules were formed. Neither fragments nor circular single-stranded Φ X174 DNA formed joint molecules with linear duplex G4 DNA, whereas the perfectly homologous combination of single-stranded G4 DNA with duplex G4 DNA yielded joint molecules. Parallel observations were made with single-stranded G4 DNA and duplex Φ X174 DNA (Table 1a).

The pairing by RecA protein of circular viral single strands with linear duplex DNA is a precise reaction in which a heteroduplex joint forms preferentially at the 3' end of the complementary strand and grows in one direction away from the end^{11,14,15}. When the linear molecules are produced by restriction cleavage of circular duplex DNA, the formation of heteroduplex DNA begins at a specific sequence. Thus with regard to the pairing of Φ X174 and G4 DNA, the published nucleotide sequences reveal the degree of homology at the sites from which heteroduplex formation begins. The nucleotide sequences of the appropriate ends of linear duplex Φ X174 DNA and G4 DNA are listed in Table 2^{21,22}. A short distance from each end, there is a subterminal region in which 1/3 to 1/2 of bases do not match. Experiments reported elsewhere have shown that a density of mismatches as low as 1 base pair in 10 can retard the growth of heteroduplex joints promoted by RecA protein¹⁵. In the case of relaxed Φ X174 and G4 DNA the greater density of mismatched sequences blocks not only the enlargement of heteroduplex joints, but also the formation of any stable joint molecules by RecA protein. Superhelicity, however, overcomes the inability of RecA protein to pair these partially homologous molecules. When we reacted fragments of single-stranded Φ X174 DNA and superhelical G4 DNA, or vice versa, with RecA protein, we observed the formation of

joint molecules (Table 1b). The mixed combinations of circular single-stranded DNA with superhelical DNA (Table 1b) or of fragments with nicked circular DNA (Table 1c) did not yield joint molecules.

The joint molecules made by combinations of Φ X174 and G4 DNA had the electron microscopic appearance of D-loops (data not shown). Measurement showed that the mean lengths of D-loops in the mismatched combinations exceeded 200 base pairs, corresponding to about 2/3 to 4/5 of the lengths of D-loops made in perfectly matched combinations (Table 3). Since the median length of perfectly matched sequences in Φ X174 and G4 is 14 bases and since no perfectly homologous sequence exceeds 30 bases, the D-loops made by Φ X174 and G4 must contain many mismatched base pairs. For example

Table 4 Summary of observations on fidelity of homologous pairing promoted by RecA protein

		Type of DNA		Formation of joint molecules	Ref.
		Single stranded	Duplex		
Genetic	G4 + Φ X174	Circular	Linear	—	This paper
		Circular	Superhelical	—	12
		Fragmented	Nicked circular	—	This paper
		Fragmented	Superhelical	+	This paper
fd + M13		Circular	Linear	+	15
		Circular	Superhelical	—	Unpublished

there is a stretch of 30 homologous bases in gene D; in the 200 bases preceding and following this stretch there are 65 and 42 mismatched bases. The other 3 homologous runs of 28–30 bases are flanked by similarly large numbers of mismatched bases²².

When superhelical DNA containing D-loops is nicked, the D-loops promptly dissociate by spontaneous branch migration²⁴. We compared the dissociation of D-loops made from homologous and mixed combinations of Φ X174 and G4 DNA. D-loops made from mixed combinations dissociated more rapidly, as expected for shorter D-loops and for D-loops containing mispaired bases (data not shown).

The experiments described above show that superhelicity promotes the formation by RecA protein of heteroduplex joints that contain many mismatches. No measurable formation of D-loops occurred in 60 min at 37°C in the absence of RecA protein, even when the DNA was perfectly homologous (Fig. 1). The requirement of RecA protein for the formation of D-loops from Φ X174 plus G4 DNA is further shown by experiments in which we tried to make D-loops from Φ X174 and G4 DNA by annealing²⁵. Homologous pairs of DNA formed 21–24% D-loops in 30 min at 75°C, but the mixed combinations G4 superhelical DNA and Φ X174 single-stranded fragments, or

Table 3 The size of D-loops made by RecA protein from partially homologous fragments and superhelical DNA

Superhelical DNA		
Fragments of ssDNA	Φ X174	G4
Φ X174	340 ± 110*	220 ± 120
G4	250 ± 130	320 ± 100

D-loops were made with superhelical DNA and single-stranded fragments as for Table 1. Measurements of the lengths of D-loops by electron microscopy were made as described previously²³.

* 14–19 D-loops were measured in each sample.

vice versa, formed none. In the absence of RecA protein, we did not observe the spontaneous formation of D-loops by Φ X174 superhelical DNA and fragments of G4 DNA at any temperature between 37° and 75 °C (Fig. 1).

Thus observations on the pairing of DNA from related small DNA phages have shown that the basis for discriminating a high degree of relatedness from perfect or nearly perfect homology exists in the mechanism by which RecA protein promotes homologous pairing (summarized in Table 4). We can begin to relate the effects of mispairing to the two steps in the reaction, namely synapsis and heteroduplex formation. The sequences of fd and M13 differ by 3% of bases; whereas Φ X174 and G4 differ by 33% of bases in the coding regions which comprise 95–96% of their genomes^{19–22}. These different degrees of potential mismatching have distinguishable effects. According to previous observations on the joint action of RecA protein and *E. coli* topoisomerase I, the imperfect homology of Φ X174 and G4 DNA interferes with synapsis, a step that precedes the formation of an extensive heteroduplex joint¹². However, if the single-stranded DNA is fragmented, and the duplex DNA is superhelical, then even 33% nonhomology does not bar the formation of joint molecules, although the frequency is reduced relative to completely homologous molecules. The 3% nonhomology of fd relative to M13 permits the formation of joint molecules in non-superhelical DNA, but interferes with the growth of heteroduplex joints that is actively promoted by RecA protein¹⁵. Nonetheless, many mismatched base pairs are included in the joint molecules formed by fd and M13 DNA, and correspondingly more mismatched base pairs are included in the joint molecules that are formed by Φ X174 and G4 DNA when superhelicity favours the reaction.

Although spontaneous uptake of homologous single-stranded fragments by superhelical DNA is driven by the energy of superhelix formation, additional energy in the form of heat is required to make the reaction go at readily measured rates. On several grounds, we suggested that this energy unwinds the helix at potential nucleation sites and that this unwinding is the rate-limiting step. Following nucleation, the superhelix energy causes the D-loop to grow because it stabilizes the product, which is a less constrained structure than the original superhelix²⁵. In the present experiments, the failure of Φ X174 and G4 DNA to form D-loops in the thermal reaction is presumably due to lack of nucleation of imperfectly matched sequences at high temperatures, in other words to instability of a nascent heteroduplex joint. RecA protein may promote the pairing of G4 and Φ X174 because it unwinds the helix at 37 °C (ref. 26) at which temperature even imperfectly paired sequences may form a nucleus that can be enlarged by the energy of superhelix formation. According to this interpretation, RecA protein, using the energy of ATP, replaces the role of thermal energy in unwinding the helix so that nucleation can occur. In the absence of superhelicity, RecA protein fails to pair Φ X174 and G4 DNA because there is a net loss of perfectly paired bases without the offsetting entropic gain that is associated with formation of a D-loop in superhelical DNA²⁵.

With regard to recombination, and possibly mutagenesis, the significant observation reported here is that RecA protein, using superhelical DNA as the recipient of a strand transfer, can make heteroduplex joints containing many mismatches.

We thank Lynn Osber for technical assistance and John Flory for helpful comments on the manuscript. This research was sponsored by grant NP 90G from the American Cancer Society and grant USPHS CA 16038–07 from the NCI.

Received 10 August; accepted 9 November 1981.

1. Fox, M. S. A. *Rev. Genet.* **12**, 47–68 (1978).
2. Radding, C. M. A. *Rev. Biochem.* **47**, 847–880 (1978).
3. Stadler, D. R. A. *Rev. Genet.* **7**, 113–127 (1973).
4. Fogel, S., Mortimer, R., Lusnak, K. & Tavares, F. *Cold Spring Harb. Symp. quant. Biol.* **43**, 1325–1341 (1978).
5. Rossignol, J.-L., Paquette, N. & Nicolas, A. *Cold Spring Harb. Symp. quant. Biol.* **43**, 1343–1352 (1978).
6. Magni, G. E. & Von Borstel, R. C. *Genetics* **47**, 1097–1108 (1962).
7. Magni, G. E. *Proc. natn. Acad. Sci. U.S.A.* **50**, 975–980 (1963).
8. Strigini, P. *Genetics* **52**, 759–776 (1965).
9. Clark, A. J. A. *Rev. Genet.* **7**, 67–86 (1973).

10. Kobayashi, I. & Ikeda, H. *Molec. gen. Genet.* **166**, 25–29 (1978).
11. DasGupta, C., Shibata, T., Cunningham, R. P. & Radding, C. M. *Cell* **22**, 437–446 (1980).
12. Cunningham, R. P., Wu, A. M., Shibata, T., DasGupta, C. & Radding, C. M. *Cell* **24**, 213–223 (1981).
13. DasGupta, C., Wu, A. M., Kahn, R., Cunningham, R. P. & Radding, C. M. *Cell* **25**, 507–516 (1981).
14. Kahn, R., Cunningham, R. P., DasGupta, C. & Radding, C. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4786–4790 (1981).
15. DasGupta, C. & Radding, C. M. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
16. Cox, M. M. & Lehman, I. R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3433–3437 (1981); *Proc. natn. Acad. Sci. U.S.A.* **78**, 6018–6022 (1981).
17. Shibata, T., DasGupta, C., Cunningham, R. P. & Radding, C. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1638–1642 (1979).
18. McEntee, K., Weinstock, G. M. & Lehman, I. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2615–2619 (1979).
19. Beck, E. *et al. Nucleic Acids Res.* **5**, 4495–4503 (1978).
20. van Wezenbeek, P. M. G. F., Hulsebos, T. J. M. & Schoenmakers, J. G. G. *Gene* **11**, 129–148 (1980).
21. Godson, G. N., Fiddes, J. C., Barrell, B. G. & Sanger, F. in *The Single-Stranded DNA Phages* (eds Denhardt, D. T., Dressler, D. & Ray, D. S.) 51–86 (Cold Spring Harbor Laboratory, New York, 1978).
22. Godson, G. N. in *The Single-Stranded DNA Phages* (eds Denhardt, D. T., Dressler, D. & Ray, D. S.) 671–695 (Cold Spring Harbor Laboratory, New York, 1978).
23. Cunningham, R. P., DasGupta, C., Shibata, T. & Radding, C. M. *Cell* **20**, 223–235 (1980).
24. Radding, C. M., Beattie, K. L., Holloman, W. K. & Wiegand, R. C. *J. molec. Biol.* **116**, 825–839 (1977).
25. Beattie, K. L., Wiegand, R. C. & Radding, C. M. *J. molec. Biol.* **116**, 783–803 (1977).
26. Cunningham, R. P., Shibata, T., DasGupta, C. & Radding, C. M. *Nature* **281**, 191–195 (1979); **282**, 426 (1979).

Three classes of mouse H-2 messenger RNA distinguished by analysis of cDNA clones

David Cosman, George Khoury & Gilbert Jay

Laboratory of Molecular Virology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205, USA

The major histocompatibility complex (MHC) is composed of a set of linked genes encoding proteins involved in the host's immune response¹. A subset of these proteins are the classical transplantation antigens, which are responsible for rapid allograft rejection and are involved in the associative recognition of foreign antigens by cytotoxic T cells^{1,2}. Three different genetic loci coding for the transplantation antigens in humans (HLA-A, -B and -C) and in mice (H-2K, D and L) have been identified^{3,4}; they are highly polymorphic⁵. The recent molecular cloning of HLA^{5,6} and H-2^{7,8} cDNA sequences provide the basis for detailed characterization of the genetic organization and expression of the MHC. We report here the identification, by analysis of cDNA clones, of at least three distinct classes of H-2 messenger RNAs that can be differentiated on the basis of their disparate 3'-non-translated regions, and describe an unusual sequence arrangement in one class of H-2-like molecule.

An adult liver cDNA library, derived from SWR/J mice (*q* haplotype) and cloned into the *Pst*I site of pBR322 (ref. 9), was screened by hybridization using as a DNA probe the insert from the HLA cDNA clone of Sood *et al.*⁶. Approximately 150,000 colonies were screened and 30 clones isolated. Excision of the DNA inserts from each of the putative H-2 clones, by digestion with *Pst*I, generated multiple fragments which show extensive overlaps between the clones (Fig. 2a), suggesting that most of the clones are related to each other.

Clone pH2 was chosen for further analysis. The entire DNA insert at the *Pst*I site of pBR322 is ~1,150 base pairs (bp) long and contains three internal *Pst*I cleavage sites (Fig. 1). The order of the four *Pst*I fragments is C (220 bp), B (250 bp), A (530 bp) and D (170 bp). Nucleotide sequence analysis of each fragment has shown that pH2 is highly homologous to the H-2 cDNA clones isolated by others^{7,8} and to the known amino acid sequences of H-2 transplantation antigens⁴. The entire pH2 insert extends from the codon for amino acid 86 (numbered according to the sequence of H-2K⁹) to the 3' end of the mRNA, including the presumptive polyadenylation signal, AAUAAA, located 11 nucleotides from the poly(A) tract (data not shown). The sequences determined corresponded to amino acids 86–138 from within fragment C, 198–236 from fragment B and 237–315 from fragment A.

Clone pH2 was used as a prototype with which to compare the other cDNA clones. Individual *Pst*I fragments of pH2 DNA were ³²P-labelled by nick-translation and hybridized to Southern blots of *Pst*I digests of DNA from the other recombinant plasmids. Figure 1 shows the maps for nine of the cDNA clones obtained by this cross-hybridization. During the mapping experiments, we noted many similarities between the clones. Most of the clones have either a 530- (pH2, 16, 18) or 570- (pH8, 12, 19, 6) bp cross-hybridizing A fragment (Fig. 2b), and a common 250-bp B fragment. The A and B fragments are internal fragments in pH2 (see Fig. 1). The clones differ extensively in the size of their 5' fragments, a presumed artefact of cDNA synthesis, as well as in their 3' D fragments. In addition, the D fragment of pH2, which contains entirely 3'-non-coding sequences immediately adjacent to the poly(A) tract, did not cross-react to any extent with the 3'-non-coding D fragments of many of the other clones.

To determine whether the 3'-non-coding regions of the various clones may be useful for distinguishing different classes of H-2 molecule, the *Pst*I D fragment (170 bp) from pH2, the D₁ fragment (185 bp) from pH8 and the D₂ fragment (135 bp) from pH12 were purified, labelled by nick-translation and hybridized to Southern blots of *Pst*I digests of the same set of nine plasmid DNAs (Fig. 2c-e). Figure 2c shows that the D fragments of clones pH2, pH16 and pH18 are indistinguishable on the basis of size or extent of cross-hybridization. The pH2 D fragment did not cross-react with any of the other clones. The pH8 D₁ fragment hybridized to itself (Fig. 2d) and pH25, pH26 and pH35 (data not shown). On the other hand, the pH12 D₂ fragment hybridized to a similar-sized fragment from pH19 (Fig. 2e) and to pH24 and pH31 (not shown). Thus at least three classes of H-2 mRNA molecule can be defined by differences in their 3'-non-coding regions, represented by clones pH2 (pH16, pH18), pH8 (pH25, pH26, pH35) and pH12 (pH19, pH24, pH31). Although clones pH5, pH6 and pH7 cannot be classified on the basis of this analysis because they lack a D fragment, they are not in the same class as pH2 (see below and Fig. 2f).

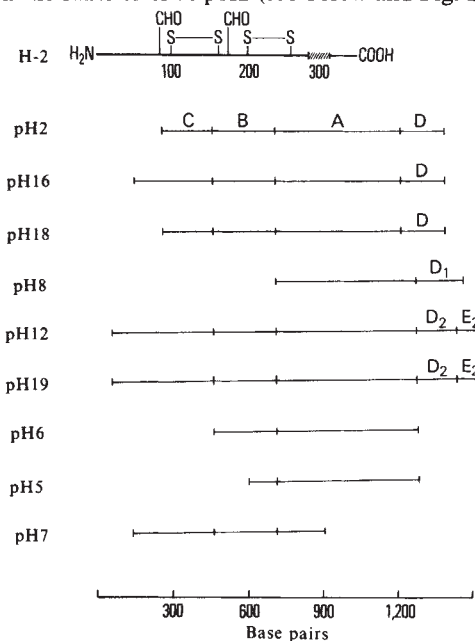


Fig. 1 The *Pst*I restriction maps of nine H-2 cDNA clones. As the recombinant plasmids were constructed by the dG-dC tailing method⁹, the cDNA inserts could be excised using *Pst*I. The multiple restriction fragments generated by *Pst*I cleavage were ordered by analysis of partial digestion products. The orientation and location of each fragment derived from clone pH2 DNA relative to the linear structure of the H-2 antigen were determined by DNA sequence analysis. The map for each of the nine clones is presented in alignment with a schematic representation of the linear structure of the H-2 antigen, including the locations of the carbohydrate side chains (CHO), the disulphide bridges (S-S) and the membrane-binding region (hatched).

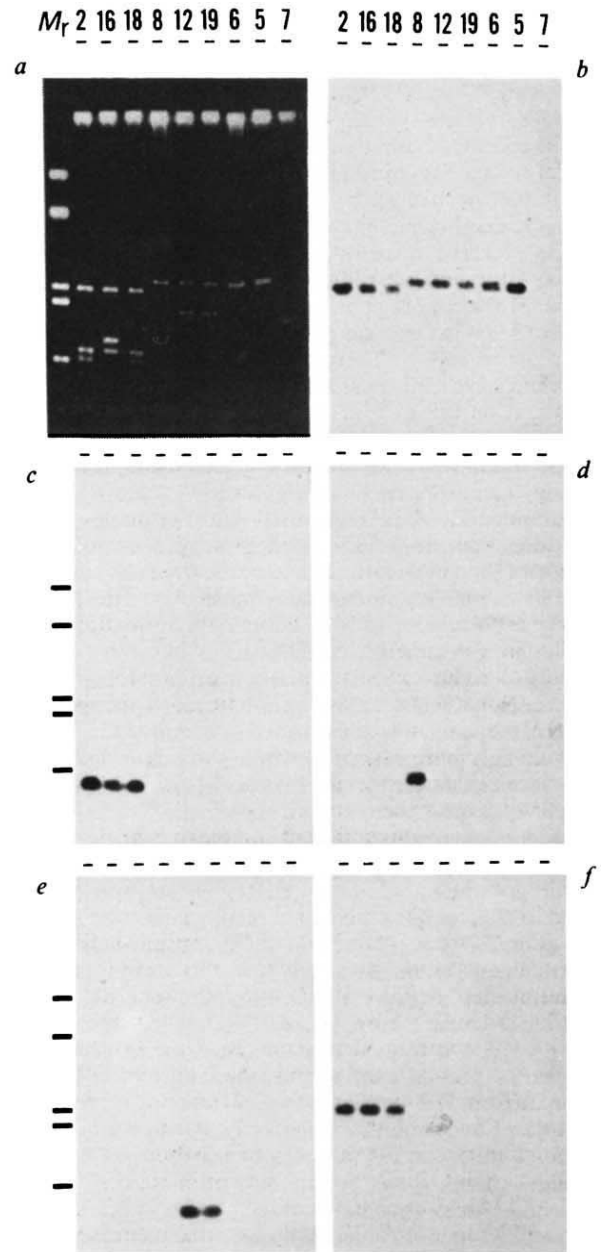
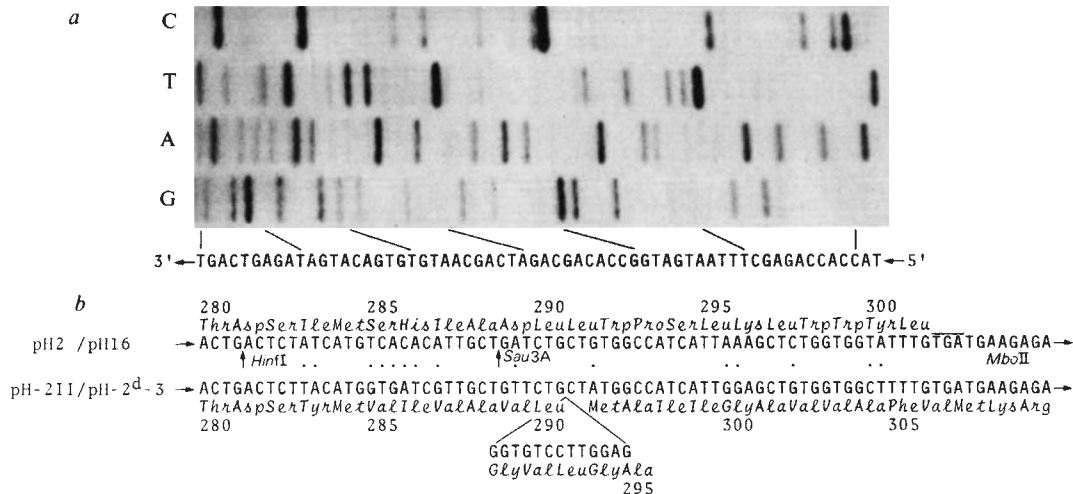


Fig. 2 Identification of different classes of H-2 cDNA clones by cross-hybridization analyses. Nine plasmid DNAs were digested with *Pst*I and fractionated by electrophoresis on 1.6% agarose gels (a). The DNA fragments were transferred from the gels to nitrocellulose sheets by the method of Southern¹². The sheets were prehybridized for 2-3 h at 65°C in 3×SSC (1×SSC = 0.15 M NaCl, 0.015 M Na citrate), 5×Denhardt's (1×Denhardt's = 0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin), 0.1% SDS, before hybridization for 16 h in the same conditions with one of several ³²P-labelled DNA probes (b-f). DNA fragments used as probes were generated by digestion with the appropriate restriction enzymes and fractionated by electrophoresis through 5-8% polyacrylamide gels. The appropriate DNA bands were eluted and ³²P-labelled by nick-translation¹³ to a specific activity of 0.5-2×10⁸ c.p.m. per μg. The 23-bp DNA fragment (f) was self-ligated with T4 DNA ligase before nick-translation. Following hybridization, the nitrocellulose sheets were washed extensively in 3×SSC, 0.1% SDS at 65°C, dried and exposed to Kodak XAR-5 films. The molecular weight markers (*M_r*), derived from simian virus 40 DNA by *Hind*III digestion, were 1,768, 1,169/1,118, 526, 447 and 215 bp long. a Is a picture of an ethidium bromide-stained gel; autoradiograms of Southern blots were obtained with the following DNA probes: b, pH2 *Pst*I A fragment; c, pH2 *Pst*I D fragment; d, pH8 *Pst*I D₁ fragment; e, pH12 *Pst*I D₂ fragment; and f, the 23-bp *Hind*III-*Sau*3A subfragment from the pH2 *Pst*I A fragment.

Fig. 3 Comparison of the amino acid sequence around the transmembrane region of the classical H-2 antigen with the predicted amino acid sequence of the corresponding region from clone pH2. *a*, Autoradiogram of a sequencing gel showing the nucleotide sequence of the antisense strand of the pH2 DNA around amino acids 280–300. The *Pst*I A fragment of pH2 DNA was cleaved with *Mbo*II and end-labelled with [γ - 32 P]ATP by the exchange reaction¹⁴. A 220-bp fragment was purified by polyacrylamide gel electrophoresis and recleaved with *Rsa*I to give two fragments. The larger of these was subjected to sequence determination by a modification of the Maat-Smith procedure^{15,16}. This sequence has been confirmed by analysis of both strands using the Maxam-Gilbert procedure¹⁴. *b*, Comparison of the pH2 sequence with the corresponding sequence from two previous isolates, pH-2II (ref. 8) and pH-2^d-3 (ref. 10). The sequences are aligned for maximum homology showing the 13-bp deletion of pH2 relative to pH-2II and pH-2^d-3. All differences in nucleotide sequences are indicated by ●. The numbers correspond to the amino acid sequence of H-2K^b (ref. 4).



The pH2 class can be further distinguished from the pH8 and pH12 classes by its slightly faster-migrating A fragment (Fig. 2a). DNA sequence analysis of pH2 showed an unusual feature in the part of the A fragment expected to encode the transmembrane portion of the H-2 molecule, from amino acids 284–307 (for reviews on the structure of the H-2 molecule, see refs 3, 4). Comparison of the sequence of the presumed transmembrane region, which is comprised of 25 uncharged and nonpolar amino acids, from two previous isolates from other laboratories, pH-2II (ref. 8) and pH-2^d-3 (ref. 10), showed perfect conservation of hydrophobic residues (Fig. 3). Analysis of the corresponding region in pH2 indicated that (1) there are extensive nucleotide and amino acid substitutions at the beginning of the transmembrane region, between positions 284 and 290, with three out of six hydrophobic residues being replaced by charged or polar amino acids; (2) this is followed by a deletion of 13 bp, which removes the codons for amino acids 291–294, located in the middle of the transmembrane region, and also the first base of the codon for amino acid 295, resulting in a shift in the reading frame; and (3) this frameshift results in the introduction of a termination codon 34 nucleotides downstream at amino acid position 302. Figure 3 shows the nucleotide sequence and predicted amino acid sequence around this region of the pH2 molecule. Note that three charged amino acids (His 286, Asp 289, Lys 296), as well as three polar residues (Ser 285, Ser 294, Pro 293), are now found interspersed throughout that region of the pH2 molecule which is equivalent to the hydrophobic transmembrane region of the classical H-2 antigen.

To determine whether this unusual sequence was present in any other cDNA clones, we purified from pH2 the 23-bp *Hinf*I–*Sau*3A restriction fragment that spans amino acids 281–288. Following self-ligation and 32 P-labelling by nick-translation, it was hybridized to Southern blots of *Pst*I-restricted plasmid DNAs (Fig. 1f). Only those clones representing the pH2 class, and containing a smaller 530 bp, rather than 570 bp, A fragment, hybridized strongly to this probe. These clones (pH2, pH16 and pH18) also have indistinguishable D fragments (Fig. 2c). Direct sequence analysis of pH16 further confirmed its identity to pH2 in this region of the molecule. This unusual pH2-like sequence is not a general feature of the transplantation antigens present in liver cells of mice of the *q* haplotype. In fact, clone pH7 has been found to have a classical hydrophobic transmembrane sequence from amino acids 280 to 293, at which point the clone is terminated by dG-dC tails (Fig. 1). The very weak hybridization of the 23-bp pH2-specific probe to pH8, -12, -19, -6, -5 and -7 (Fig. 2f) is probably due to the homology between amino acids 281 and 284 (see Fig. 3).

The 250-nucleotide sequence immediately upstream from this unusual region of pH2 has been determined and contains neither in-frame nonsense codons nor frameshifts (data not shown). This region, between amino acids 198 and 279, shows 88% homology to the H-2K^b amino acid sequence⁴. The presence of this unusual sequence in liver mRNA raises the possibility that the mRNA represented by pH2 encodes a shortened form of the H-2 antigen (301 amino acids) that might have a different biological function. Such a truncated antigen is expected to contain the entire biologically active portion of the H-2 molecule. The absence of a hydrophobic region in this putative polypeptide makes it unlikely that it could be inserted into the plasma membrane. The loss of the C-terminal hydrophilic tail would preclude it from interacting with cytoplasmic components. It is possible that such a molecule could be secreted.

Although an altered form of the H-2 antigen has not previously been described, we have observed a protein of molecular weight 35,000 that can be immunoprecipitated by anti-H-2 antiserum from cell-free translation products of liver mRNA (unpublished results). It remains to be determined whether this protein is the product of the mRNA represented by pH2. Until a protein product of this mRNA is identified, we cannot exclude the possibility that pH2 represents a transcriptionally active H-2-like pseudogene.

This study demonstrates that different classes of H-2 cDNA clones can be identified on the basis of non-homologous 3'-untranslated sequences. The classes cannot yet be assigned to the H-2K, -D or -L loci as complete amino acid sequences from the *q* haplotype are not available. This observation should provide both a method for differentiating between members of a multi-gene family and a clearer understanding of their genomic organization and expression.

After this manuscript had been submitted, Steinmetz *et al.*¹¹ published the sequence of an H-2-like pseudogene with unusual changes in the transmembrane region that are distinct from those in pH2. It was not determined whether this pseudogene was transcribed.

We acknowledge the generous gifts of the HLA-B cDNA plasmid from Dr Sherman Weissman and the mouse liver cDNA library from Drs John Monahan and Michael Tocci.

Received 25 August; accepted 4 November 1981.

1. Klein, J. *Science* **203**, 516–521 (1979).
2. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 51–177 (1979).
3. Ploegh, H. L., Orr, H. T. & Strominger, J. L. *Cell* **24**, 287–299 (1981).
4. Nathanson, S. G. *et al.* *Rev. Biochem.* **50**, 1025–1052 (1981).
5. Ploegh, H. L., Orr, H. T. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6081–6085 (1980).

6. Sood, A. K., Pereira, D. & Weissman, S. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 616–620 (1981).
7. Kvist, S. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **78**, 2772–2776 (1981).
8. Steinmetz, M. *et al.* *Cell* **24**, 125–134 (1981).
9. Norgard, M. V., Tocci, M. J. & Monahan, J. J. *J. biol. Chem.* **255**, 7665–7672 (1980).
10. Bregegere, F. *et al.* *Nature* **292**, 78–81 (1981).
11. Steinmetz, M. *et al.* *Cell* **25**, 683–692 (1981).
12. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
13. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).
14. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
15. Maat, J. & Smith, A. J. H. *Nucleic Acids Res.* **5**, 4537–4545 (1978).
16. Seif, I., Khoury, G. & Dhar, R. *Nucleic Acids Res.* **8**, 2225–2240 (1980).

Cloning and analysis of cDNAs encoding plant storage protein precursors

Ronald R. D. Croy, Grantley W. Lycett,
John A. Gatehouse, Jennifer N. Yarwood
& Donald Boulter

Department of Botany, University of Durham, Science Laboratories,
South Road, Durham DH1 3LE, UK

The major storage protein in many legume seeds is legumin, a hexameric molecule of molecular weight (M_r) 360,000–400,000 consisting of six subunit pairs, each of which comprises a 40,000- M_r , acidic polypeptide linked by disulphide bonds to a 20,000- M_r , basic polypeptide¹. We have recently shown that these subunit pairs are synthesized, *in vivo* and *in vitro* in *Pisum* and *Vicia*^{2,3}, as a 60,000- M_r precursor polypeptide, putatively containing one copy each of the acidic and basic subunits covalently linked together. We describe here the cloning and identification of cDNA sequences coding for the legumin precursor. Additionally, we have used these cDNAs to verify that the legumin mRNAs are of sufficient size to code for the precursor and to demonstrate the presence of four single-copy legumin genes in genomic DNA. Possible mechanisms for the processing of the legumin precursor are discussed. Identification of cloned cDNA sequences coding for vicilin, the secondary storage protein of *Pisum sativum*⁴, is also reported.

Double-stranded DNA complementary to poly(A)⁺ RNA from developing seed cotyledons was prepared^{4,5} and cloned in plasmid pBR322^{6,7} as described in Fig. 1 legend. Transformed colonies strongly hybridizing to mRNA or cDNA probes were selected as putative legumin and vicilin cDNA clones and plasmid 'minipreps' made according to Klein *et al.*⁷. Southern 'blots' of cDNA inserts, excised from the recombinant plasmids, strongly hybridized mRNA and cDNA probes, further confirming the presence of cDNA inserts. The cDNA inserts were used to prepare and identify complementary mRNAs by hybrid-selected translation^{8,9} (Fig. 1).

An initial selection of 12 clones contained 6 cDNAs encoding vicilin 50,000 and 47,000- M_r subunits^{3,10–13}, two encoding the legumin 60,000- M_r precursor^{2,3} and several others coding for unidentified polypeptide products (Fig. 1).

Physical maps of two legumin cDNAs have been constructed from restriction endonuclease analyses and are shown in Fig. 2 together with the sequencing strategies. The legumin cDNA, pRC 2.2.4, identified by hybrid-selected translation, has been completely sequenced and corresponds to a 638-nucleotide sequence from the 3' end of the legumin mRNA (Fig. 3). There is, in addition, a polyadenylated sequence at the 3' terminus. The second legumin cDNA (pRC 2.11.7), selected by screening the cDNA clone bank with pRC 2.2.4 cDNA, has been partially sequenced. It is essentially identical to 2.2.4, but is larger by about 150 nucleotides extending towards the 5' end of the message.

The close agreement of the predicted amino acid composition, tryptic peptide and N-terminal sequence data with published data^{1,14,15} unequivocally identifies the nucleotide sequence as that of a legumin cDNA and also locates the basic (20,000 M_r) subunit at the 3' end of the legumin messenger coding sequence. The additional coding sequence of pRC 2.11.7 shows that the

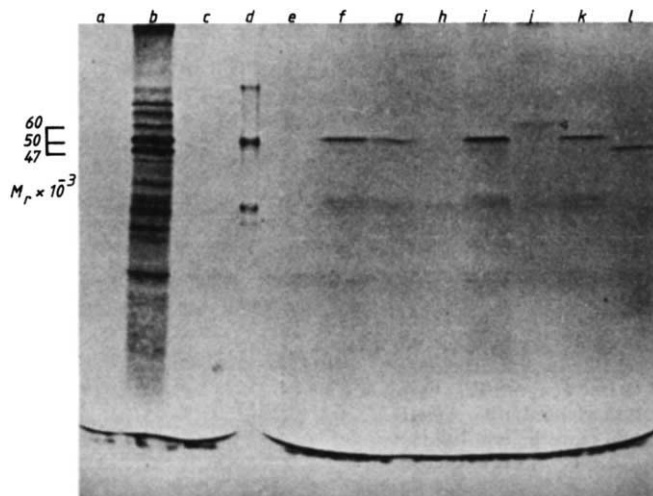


Fig. 1 Identification of recombinant plasmids containing cDNAs by translation of hybrid-selected mRNAs. Polyadenylated RNA, purified from 14-day-old pea cotyledons, was used to synthesize double-stranded cDNA by avian myeloblastosis virus reverse transcriptase and DNA polymerase I as described previously^{4,5}. The double-stranded cDNA was cloned in the *Bam*HI site of pBR322 after attachment of *Bam*HI synthetic oligonucleotide linkers, essentially by the method of Sippel *et al.*⁶. 1 μ g of ³²P- or ³H-labelled cDNA, blunt-ended with DNA polymerase I, was ligated to 400 pmol of 5'-phosphorylated linkers for 15 h at 12.5 °C (ref. 6). After digestion with an excess of *Bam*HI, the cDNA was fractionated on a Sepharose 6B column (100 mm $\times$ 15 mm diameter) and the fractions containing the largest cDNAs were pooled and recovered by alcohol precipitation. pBR322 (5 μ g) linearized with *Bam*HI, was treated with alkaline phosphatase to remove the 5'-phosphate groups (to prevent self-ligation and non-recombinant transformations^{6,44}) and after extensive phenol extraction the plasmid was passed through a Sephadex G-50 column and recovered by alcohol precipitation. The cDNA was ligated to the plasmid for 48 h at 12.5 °C. 0.5–1.0 μ g of the recombinant DNA was used to transform competent cells prepared from RecA⁺ or BedBC⁺ strains of *Escherichia coli* K-12, after 24 h incubation in calcium chloride to enhance competence²². Transformants were selected on ampicillin plates and screened by colony hybridization on nitrocellulose filters with ³²P-labelled cDNA and mRNA⁴⁵. Hybridizations were performed in the presence of polyadenylic acid to prevent any non-specific binding of probes to poly(dT-dA) regions in the clones⁴⁴. Probing with pea ribosomal RNA showed the complete absence of any ribosomal cDNA clones. cDNA containing plasmids, purified by the method of Clewell^{4,6} and Katz *et al.*⁴⁷, were restricted with *Bam*HI to excise the cDNA inserts, and 20–30 μ g aliquots of total DNA were denatured and applied to freshly prepared diazobenzyloxymethyl paper⁹ as described by Smith *et al.*⁸. Poly(A)⁺ mRNA was allowed to hybridize, 5 μ g per filter, to the plasmids and specifically bound mRNAs were eluted⁸ and translated in the reticulocyte lysate system³. Translation products labelled with ³H-leucine were separated and visualized by SDS-polyacrylamide gel electrophoresis followed by fluorography⁴⁸. Track b shows the translation products from total poly(A) mRNA: tracks a and c show, respectively, endogenous translation products from RNA nonspecifically bound to pBR322. Tracks e–l show the translation products from mRNAs specifically bound to cDNA plasmids. pRC 2.1.7, 2.1.8, 2.1.18, 2.1.20, 2.2.1, 2.2.4 (band indicated with an open arrow), 2.2.5, 2.2.10, respectively. (Track d shows a ³H-labelled standard protein preparation.) Legumin 60,000 M_r precursor and vicilin 50,000 and 47,000 M_r precursors were identified as described previously^{3,4}. Characterized and ³²P-nick translated²⁸ cDNAs, prepared as described in Fig. 4 legend, were used to rescreen the cDNA clone bank for related but longer cDNA sequences. Clones were arrayed and grown on nitrocellulose filters and subjected to colony hybridization using the legumin cDNA insert from pRC 2.2.4. In this way the complementary and larger legumin cDNA, pRC 2.11.7, was detected and isolated.

amino acid sequence extends unbroken upstream beyond the N terminus of the basic subunit for at least 30 amino acids, presumably past or approaching the C terminus of the acidic subunit^{1,16}. There are no termination or initiation codons in this pre-N-terminal sequence, confirming the legumin precursor theory³. The sequence data also show that the legumin mRNA has a significant non-coding sequence between the coding sequence and the poly(A) tail. This 137-nucleotide sequence is A + T rich, in common with many other eukaryotic mRNAs, and contains the hexanucleotide sequence -AATAAA-, about 20 nucleotides upstream from the 3' end, as also found in other eukaryotic mRNAs^{17–22}.

As further evidence for synthesis of legumin as a single polypeptide precursor, we have determined the size of the

legumin mRNA. Total and polyadenylated RNA preparations from developing cotyledons and leaves were denatured by glyoxalation²³ and size fractionated by agarose electrophoresis, and after transfer to nitrocellulose paper²⁴, the 'Northern' blots were probed with ³²P-labelled legumin cDNAs (Fig. 4). Legumin cDNA hybridized strongly to a single component in both seed RNA preparations corresponding to a 19S mRNA (2.2×10^3 bases), sufficiently long to code for a precursor polypeptide of at least 60,000 M_r. Although hybridization of legumin cDNA to smaller mRNA species was observed in seed polyadenylated RNA (Fig. 4), the binding was never to discrete mRNA sizes and was unique to polyadenylated preparations. No such hybridization was observed in total seed RNA preparations, indicating the absence of small mRNA species which may have coded for the smaller (basic) polypeptide components found in legumin isolated from mature seeds. We conclude from these results that extensive degradation of the 19S legumin mRNA occurs during the preparation of polyadenylated RNA, in agreement with the reports of Dunham *et al.*²⁵. There was also no hybridization of the probes to any components in the leaf polyadenylated RNA preparations, verifying the tissue-specific expression of the storage protein genes²⁶.

We have also used the legumin cDNAs to probe for the homologous sequences in genomic DNA. Pea genomic DNA was prepared from leaf tissue²⁷ and equal aliquots were digested to completion with *Eco*RI or *Hind*III restriction endonucleases (which do not cut within the cDNA sequence) (Fig. 5). After separation on agarose gels, the genomic DNA fragments were blotted on to nitrocellulose paper²⁴ and then hybridized to an excess of ³²P-nick-translated²⁸ cDNA insert (specific activity $\geq 2 \times 10^7$ c.p.m. per μ g) from pRC 2.2.4 (Fig. 5). Various amounts of linearized pRC 2.2.4 corresponding to different gene copy numbers were also included on the gels²⁹ (Fig. 5). We have only been able to detect four *Eco*RI genomic fragments (excluding any intron sites) hybridizing to legumin cDNA and an estimate of the intensities of hybridization to genomic fragments and cDNA plasmids indicates a single gene copy per fragment and therefore an apparent maximum complement of four legumin genes per haploid genome. This agrees with the recent reports of Spencer and Higgins³⁰ of a set of at least three legumin precursors found *in vivo*, and of J.A.G. *et al.* (unpublished) of at least three legumin precursors *in vitro*. Other reports have indicated that up to four major legumin species exist in mature pea seeds, depending on the genetic line^{16,31-34}. However, there have also been many reports of a greater number of minor legumin species in pea and other legumes, based on polypeptide

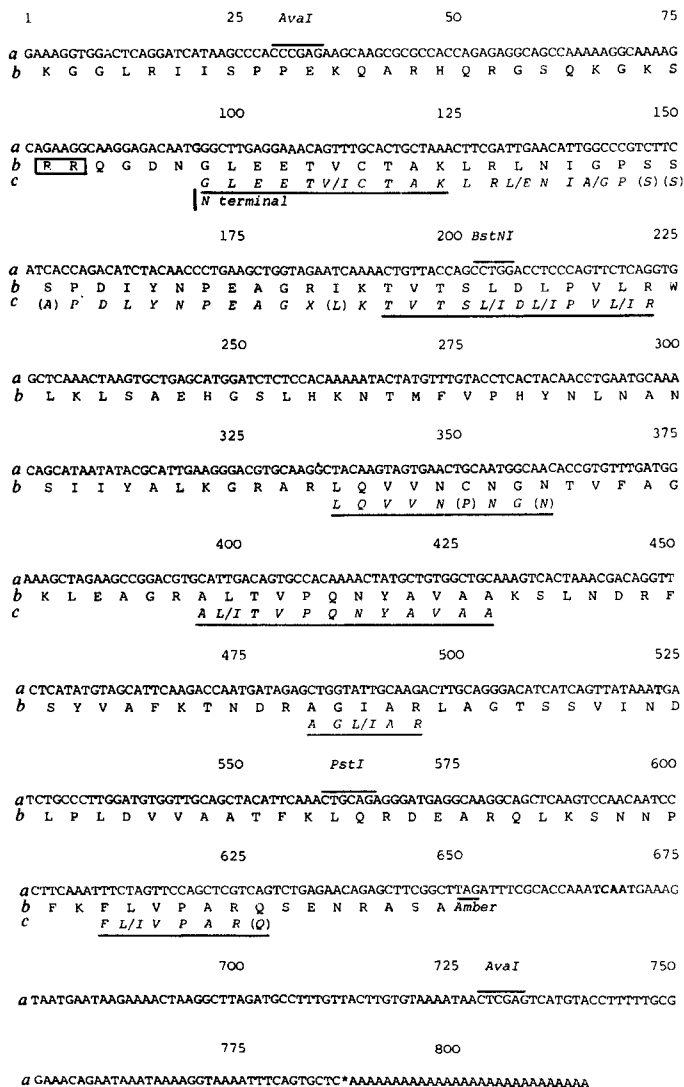


Fig. 3 Composite cDNA sequence coding for the legumin basic subunit. The figure shows the DNA (mRNA) sequence (line a) derived from pRC 2.11.7 and pRC 2.2.4. Numbering in the nucleotide sequence corresponds to that of pRC 2.11.7 in Fig. 3. The asterisk at position 787 indicates the only position at which the two sequences differ, with an adenine residue occurring in 2.2.4 and a thymine residue in 2.11.7. Restriction endonuclease sites are overlined and the termination codon is underlined, as are the overlapping AATAAA hexanucleotide sequences at the 3' end of the mRNA. DNA sequencing was by the methods of Seif *et al.*⁴⁹ or Maxam and Gilbert⁵⁰. Where the method of Seif *et al.*⁴⁹ was used, the products of forward and backward reactions were pooled before lyophilization. In both cases, end-labelled DNA fragments were prepared according to Maxam and Gilbert and sequencing gels (6 and 8% polyacrylamide) were prepared as described by Smith⁵¹ and run at 1,500 V. The predicted amino acid sequence (line b) was obtained from the DNA sequence by the use of a computer program (provided by Dr M. D. Watson). Line c shows peptide amino acid sequences matching the predicted sequence. The N-terminal peptide sequence was obtained from the published data of Casey *et al.*¹⁵ and all other sequences (underlined) were obtained by sequencing purified tryptic peptides from the legumin basic polypeptides (J.N.Y., in preparation). Basic subunits were isolated from total legumin by ion-exchange chromatography in formamide, as described by Gatehouse *et al.*³⁷. The separated basic subunits were extensively dialysed against distilled water and lyophilized. Tryptic digestions were carried out for 6 h at 37 °C (100/2 w/w) and the peptide mixture lyophilized before separation by HPLC. The tryptic peptides were separated on a reverse-phase column (C18, 10 μ , Varian MCH-10) using a trifluoroacetic acid/acetonitrile gradient (0–70% acetonitrile). Separated peptides were collected and lyophilized before being sequenced by the DABITC method of Chang *et al.*⁵².

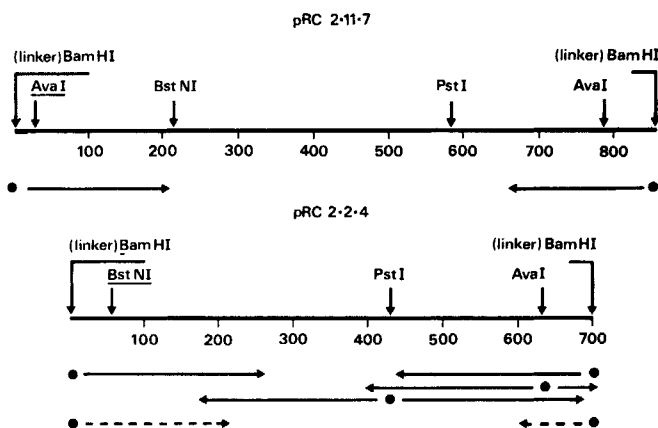


Fig. 2 Restriction maps of cDNAs from plasmids pRC 2.11.7 and pRC 2.2.4 showing the positions of restriction enzyme targets used in the sequencing experiments. The horizontal scales represent nucleotide base pairs from the 5' end of the coding strand. Numbers of nucleotide base pairs in pRC 2.11.7 correspond to the scale in the nucleotide sequence (Fig. 3). The horizontal arrows indicate the extent of the fragments sequenced and the direction of sequence reading from the 5'-labelled ends (●). Sequencing was by the method of Seif *et al.*⁴⁹ (continuous arrows) or that of Maxam and Gilbert⁵⁰ (broken arrows).

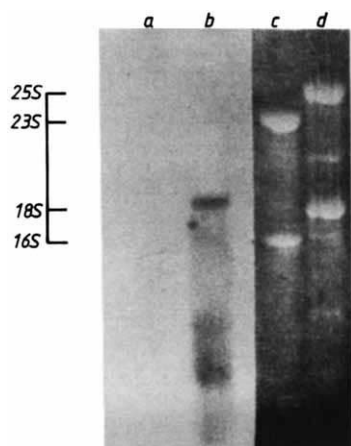


Fig. 4 Identification and sizing of legumin mRNA. Poly(A)⁺ mRNA (1 µg) or total seed RNA (5 µg), prepared according to the methods of Hall *et al.*⁵³ and Girard⁵⁴, were denatured by glyoxylation and electrophoresed on 1.5% agarose gels by the method of McMaster and Carmichael²³. The nucleic acids were transferred to nitrocellulose paper according to Thomas²⁴ and hybridized with ³²P-end-labelled or -nick-translated²⁸ cDNAs (specific activity 10⁷ c.p.m. per µg). cDNAs were excised from the plasmids with *Bam*HI and separated on 0.5% agarose gels. The cDNAs were electroeluted from gel slices into dialysis tubing and recovered by alcohol precipitation. After end labelling or nick-translation, the cDNAs were re-electrophoresed through 0.5% agarose into dialysis tubing and used without alcohol precipitation for the hybridizations. Hybridizations were performed in the presence of 100 µg ml⁻¹ polyadenylic acid to prevent nonspecific binding⁴⁴. Autoradiographs of the hybridizations are shown as follows: track a, poly(A) mRNA from leaves; track b, poly(A) mRNA from developing seeds hybridized with ³²P-legumin cDNA (pRC 2.2.4). Tracks c and d show standard ribosomal RNAs, *E. coli* 16S and 23S, and pea 18S and 25S respectively, stained with acridine orange²³.

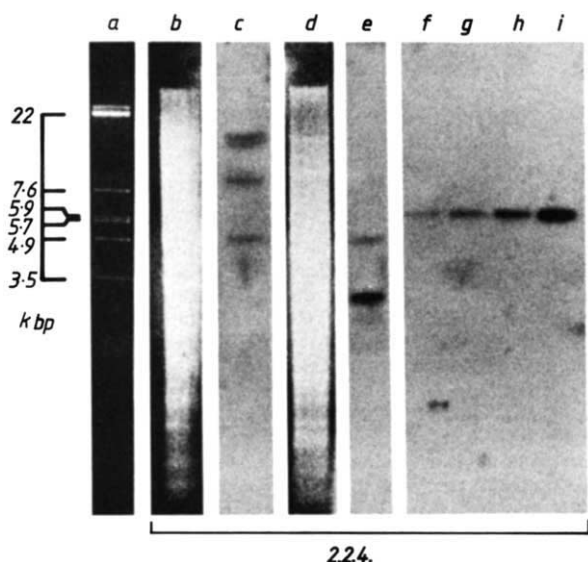


Fig. 5 Use of specific cDNAs to probe for legumin genes. Genomic DNA was isolated from developing pea leaves and purified twice on caesium chloride gradients²⁷. Aliquots (5 µg) of the pure DNA were digested to completion with *Eco*RI, or *Hind*III. The fragments were then separated by electrophoresis overnight on 0.5% agarose gels and transferred to nitrocellulose paper²⁴. The genomic blots were hybridized with ³²P-cDNAs prepared as described in Fig. 4 legend (specific activity ≥10⁷ c.p.m. per µg). Track a shows standard fragments of an *Eco*RI digest of λNM258, stained with ethidium bromide; tracks b, c show *Eco*RI and tracks d, e *Hind*III digests of pea genomic DNA stained with ethidium bromide (tracks b, d) and probed with ³²P-legumin cDNA (pRC 2.2.4) (tracks c, e); tracks f–i show 2.6, 5.3, 10.6, 26.5 pg of linearized plasmid pRC 2.2.4 (corresponding to 0.5, 1, 2 and 5 legumin gene copies in 5 µg of genomic DNA²⁷) and hybridized with ³²P-legumin cDNA.

analyses and amino acid sequencing of purified acidic and basic polypeptides (refs 16, 31–33, 35 and Nielsen, personal communication)—that is, more than could be specified by four genes (for a review see ref. 36). Some of the additional heterogeneity may be accounted for by post-translational modifications^{3,37} or, on the other hand, the probe used here may not have hybridized to all legumin DNA sequences. Similar analysis of genomic DNA isolated from pea seed cotyledons gave no evidence for selective amplification or alteration of the legumin genes in these tissues, supporting previous work³⁸.

The amino acid sequence deduced from the legumin cDNA nucleotide sequence clearly shows that the coding sequence for the acidic subunit precedes that of the basic subunit in the mRNA. The processing of legumin¹ seems to be analogous to that of animal insulin, where excision of a linking (C) peptide from the precursor proinsulin by proteolytic cleavage at Arg-Arg and Arg-Lys paired basic residues produces two disulphide-linked subunits^{39,40}. Such adjacent basic residues have a prominent role as endopeptidase recognition sites in the processing of many animal polypeptide hormones and secretory proteins^{21,40,41}. Although a similar linking peptide has not been demonstrated for the legumin precursor, examination of the deduced amino acid sequence upstream from the N terminus of the basic subunit reveals an unusually high content of the basic amino acids and the presence of an Arg-Arg sequence only five residues from the N-terminal amino acid. Although this may be a processing site, clearly further peptidolytic processing would be required to produce the N-terminal glycine. Similar 60,000-*M_r* legumin precursors have been demonstrated for *Vicia faba*³ and *Glycine max* (unpublished results), suggesting a common processing mechanism.

The relatively long half life of the legumin precursors established by *in vivo* labelling^{3,28} is in accordance with the above suggestions, and implicates a secondary membrane system distinct from the endoplasmic reticulum, such as the Golgi apparatus, in this processing. Harris⁴² has suggested that the Golgi apparatus may be involved in the post-translational modifications and internal secretion of seed proteins, and it is known that the processing of proinsulin is carried out in this membrane system.

Results to be published elsewhere suggest that, like legumin, the vicilin polypeptide precursors^{4,12} are synthesized only from high molecular weight mRNAs (17S) and these vicilin mRNAs are also derived from genes present in single or low copy number in the pea genome.

We thank Philippa Brown, Dave Bown and Ray Taylor for technical assistance, and Bill Brammar and Dick Flavell for their help and advice on cloning strategy. The financial support of the SRC is acknowledged.

Received 17 August; accepted 13 November 1981.

- Derbyshire, E., Wright, D. J. & Boulter, D. *Phytochemistry* **15**, 3–24 (1976).
- Evans, I. M. *et al. Planta* **144**, 455–462 (1979).
- Croy, R. R. D., Gatehouse, J. A., Evans, I. M. & Boulter, D. *Planta* **148**, 49–56 (1980).
- Croy, R. R. D., Gatehouse, J. A., Evans, I. M. & Boulter, D. *Planta* **148**, 57–63 (1980).
- Bolivar, R. *et al. Gene* **2**, 95–113 (1977).
- Sippel, A. E. *et al. Nucleic Acids Res.* **5**, 3275–3294 (1978).
- Klein, R. D., Selsing, E. & Wells, R. D. *Plasmid* **3**, 88–91 (1980).
- Smith, D. F., Searle, P. F. & Williams, J. G. *Nucleic Acids Res.* **6**, 487–506 (1979).
- Alwine, J. C., Kemp, D. J. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5350–5454 (1977).
- Gatehouse, J. A., Croy, R. R. D., Morton, H., Tyler, M. & Boulter, D. *Eur. J. Biochem.* **118**, 627–633 (1981).
- Thomson, J. A., Schroeder, H. E. & Dudman, W. F. *Aust. J. Pl. Physiol.* **5**, 263–279 (1978).
- Higgins, T. J. V. & Spencer, D. *Pl. Physiol.* **67**, 205–211 (1981).
- Croy, R. R. D., Gatehouse, J. A., Tyler, M. L. & Boulter, D. *Biochem. J.* **191**, 509–516 (1980).
- Croy, R. R. D., Gatehouse, J. A., Lycett, G. W. & Boulter, D. *Biochem. Soc. Trans.* **9**, 254P, S29–18 (1981).
- Casey, R., March, J. F. & Sanger, E. *Phytochemistry* **20**, 161–163 (1981).
- Croy, R. R. D., Derbyshire, E., Krishna, T. G. & Boulter, D. *New Phytol.* **83**, 29–35 (1979).
- Kafatos, F. C., Efstratiadis, A., Forget, B. G. & Weissman, S. M. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5618–5622 (1977).
- Bell, G. I. *et al. Nature* **282**, 525–527 (1979).
- Cheng, C. C. *et al. J. molec. Biol.* **107**, 527–547 (1976).
- Proudfoot, N. J. & Brownlee, G. G. *Nature* **252**, 359–362 (1974).
- Hobart, P., Crawford, R., Shen, L., Pictet, R. & Rutter, W. J. *Nature* **288**, 137–141 (1980).
- Dagert, M. & Ehrlich, S. D. *Gene* **6**, 23–28 (1979).
- McMaster, G. K. & Carmichael, G. G. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4835–4838 (1977).
- Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5202–5205 (1980).
- Dunham, S. M., Tenniswood, M. & Tata, J. R. *Biochem. Soc. Trans.* **9**, 202P, S17–18 (1981).
- Miller, A. A. *Rev. Pl. Physiol.* **26**, 53–72 (1975).

27. Graham, D. E. *Analyt. Biochem.* **85**, 609–613 (1978).
28. Amersham International *Nick Translation Kit Pamphlet* P1/86/81/4.
29. Young, R. A., Hagenbüchle, O. & Schibler, U. *Cell* **23**, 451–458 (1981).
30. Spencer, D. & Higgins, T. J. V. *Biochem. Int.* **1**, 502–509 (1980).
31. Krishna, T. G., Croy, R. R. D. & Boulter, D. *Phytochemistry* **18**, 1879–1880 (1979).
32. Casey, R. *Biochem. J.* **177**, 509–520 (1979).
33. Matta, N., Gatehouse, J. A. & Boulter, D. *J. exp. Bot.* (in the press).
34. Guldager, P. *Theor. appl. Genet.* **53**, 241–250 (1978).
35. Moreira, M. A., Hermodson, M. A., Larkins, B. A. & Nielsen, N. C. *J. biol. Chem.* **254**, 9921–9926 (1979).
36. Matta, N. thesis, Univ. Durham (1981).
37. Gatehouse, J. A., Croy, R. R. D. & Boulter, D. *Biochem. J.* **185**, 497–503 (1980).
38. Miller, A. & Whitfield, P. R. *Pl. Physiol.* **51**, 1005–1010 (1973).
39. Steiner, D. F. *Diabetes* **26**, 322–340 (1977).
40. Steiner, D. F., Quinn, P. S., Chan, S. J., Marsh, J. & Tager, H. S. *Ann. N.Y. Acad. Sci.* **343**, 1–16 (1980).
41. Cohen, S. N. *et al. Ann. N.Y. Acad. Sci.* **343**, 415–423 (1980).
42. Harris, N. *Planta* **146**, 63–69 (1979).
43. Evans, I. M., Croy, R. R. D., Brown, P. & Boulter, D. *Biochim. biophys. Acta* **610**, 81–95 (1980).
44. Bedbrook, J. R., Smith, S. M. & Ellis, R. J. *Nature* **287**, 692–697 (1980).
45. Grunstein, M. & Wallis, J. *Meth. Enzym.* **68**, 379–389 (1979).
46. Clewell, D. B. *J. Bact.* **110**, 667–676 (1972).
47. Katz, L., Williams, P. H., Sato, S., Leavitt, R. W. & Helinski, D. R. *Biochemistry* **16**, 1677–1683 (1977).
48. Bonner, W. M. & Laskey, R. A. *Eur. J. Biochem.* **46**, 83–88 (1974).
49. Seif, I., Khoury, G. & Dhar, R. *Nucleic Acids Res.* **8**, 2225–2240 (1980).
50. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 449–560 (1980).
51. Smith, A. J. H. *Meth. Enzym.* **65**, 560–580 (1980).
52. Chang, J. Y., Braver, D. & Wittam-Liebold, B. *FEBS Lett.* **93**, 205–214 (1978).
53. Hall, T. C. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3196–3200 (1978).
54. Girard, M. *Meth. Enzym.* **12A**, 581–588 (1967).

Carboxyl–carboxylate interactions in proteins

Lindsay Sawyer* & Michael N. G. James†

MRC Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7

Much of our present understanding of macromolecular architecture has come from the extrapolation to large molecules of rules derived from analysis of small molecules. This is exemplified by the work of Pauling *et al.*¹ on polypeptide conformation, which depended on the insight afforded by the crystal structures of individual amino acids and some dipeptides. Now that accurate protein crystal structures are available, it appears that most of the hydrogen-bonding patterns that are observed have already been found in the structures of small molecules. Thus we can learn much about weak interactions in proteins by analysing small molecule crystal structures. Here we comment on the existence of hydrogen bonds between two carboxyl groups, forming what amounts to the acid salts of monobasic carboxylic acids. We believe these interactions are more important in stabilizing the structures of protein crystals and multi-subunit complexes at low pH than has been realized previously. Furthermore, the interaction contributes to hydrolytic activity and may be involved in inter-subunit recognition.

Speakman² has described type A and B monobasic acid salts of general formula MHX_2 . Examples of type B, having the formula $\text{M}^+\text{X}^-\text{HX}$, rubidium hydrogen bisglycollate hydrate³ and potassium hydrogen di-*p*-nitrobenzoate⁴. On the other hand, type A salts have formally equivalent X units requiring the $\text{O}\cdots\text{H}\cdots\text{O}$ bond to be situated on a symmetry element. For example, the carboxyl groups in the crystals of potassium hydrogen bisphenyl acetate are related by a centre of symmetry⁵ and, of more interest here, sodium hydrogen diacetate has a dyad axis relating the acidic groups⁶. Only one proton is involved in the interaction so that the carboxyl groups adopt an offset arrangement which allows a variety of conformations. We have defined these conformations in terms of five angles and the distance of the hydrogen bond (see Table 1). The remaining negative charge in the case of the small molecule is balanced by a nearby cation, M. The structures of many such acid salts have been reported and it is not surprising to find that they exist in proteins. Table 1 shows some typical carboxyl group dimensions for these salts.

'Pronase', a crude extract from cultures of *Streptomyces griseus*, contains several distinct proteolytic enzymes, of which the crystal structures of protease A (SGPA) and protease B (SGPB) have been refined at high resolution. Each has a distinctive carboxyl–carboxylate hydrogen bond as an essential intermolecular interaction. SGPB crystallizes in space group P2_12_12 from 1.2 M KH_2PO_4 , pH 4.2 (ref. 7). Refinement has given an *R* factor of 14.9% for the data to 1.7 Å. From Fig. 1a, which shows the region around the crystallographic dyad, it can be seen that the Glu 137 residues (192A) form a close interaction. (Numbers in parentheses refer to alternative numbering schemes^{9,10}.) This relationship is precisely that found in type A acid salts and the separation is consistent with a very short hydrogen bond (see Table 1). Note that there is no balancing cation as observed in the small molecule structures. O255 could perhaps be considered a cation but the occupancy and temperature factor would be consistent only with H_3O^+ . However, in every case where an oxonium ion has been examined in a crystal, there is a pyramidal arrangement with three hydrogen bonds of ~ 2.6 Å (ref. 8). The hydrogen bond from Glu 137 O[−] (192A) to Cys 162 N (220) will partly counteract the extra negative charge because of the peptide dipole. At pH ~ 4 , partial protonation of a glutamic acid side chain ($\text{pK}_a \sim 4.3$) is expected. Removal of

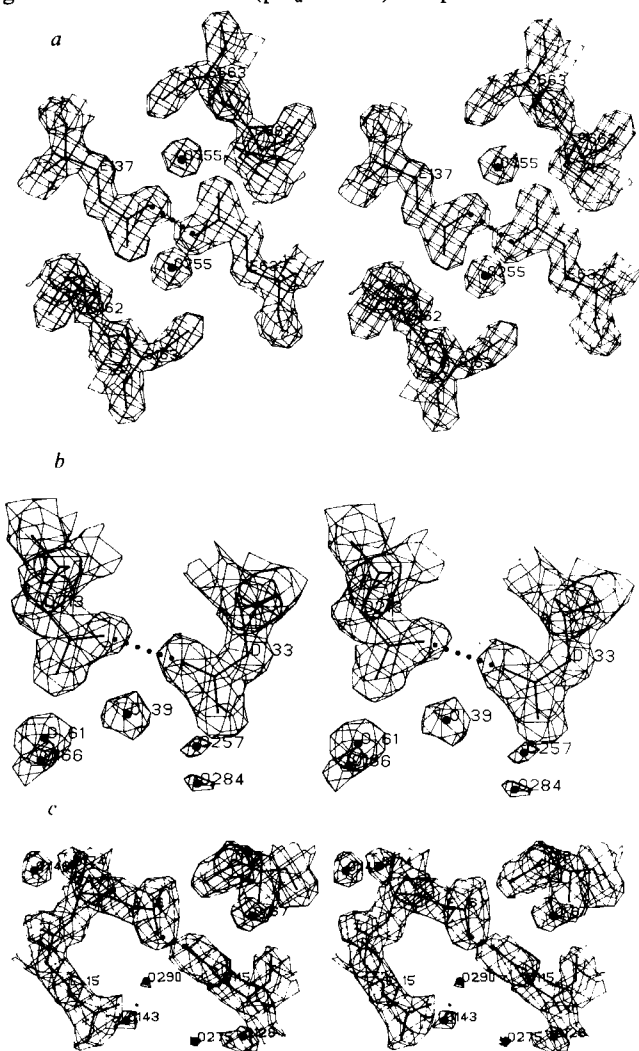


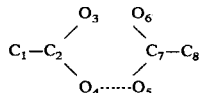
Fig. 1 The carboxyl–carboxylate interactions in *S. griseus* protease B and penicillipepsin. In each case, the hydrogen bond is shown as a dotted line but for clarity the other hydrogen bonds have been omitted. *a*, Type A interaction in SGPB viewed in the direction of the dyad axis parallel to *c*. The contour level is $0.6 \text{ e} \text{ \AA}^{-3}$. For ease of identification the residues of the twofold-related molecule have had 500 added to their sequence numbers. *b*, Type B interaction of the two catalytic residues in penicillipepsin, Asp 33(32) and Asp 213(215). The contour level is $0.5 \text{ e} \text{ \AA}^{-3}$. *c*, The second type B interaction in penicillipepsin involves Asp 115(114) and Glu 16(13). The contour level is $0.45 \text{ e} \text{ \AA}^{-3}$.

* Permanent address: Department of Chemistry, Napier College, Edinburgh EH10 5DT, UK.
† To whom correspondence should be addressed.

Table 1 Distances and angles for carboxyl-carboxylate interactions

Compound	κ_1	κ_2	κ_3	κ_4	κ_5	d (Å)	Type
Rubidium bisglycolate hydrate	172.6	172.6	57.0	115.7	120.6	2.53	B
Potassium hydrogen di- <i>p</i> -nitrobenzoate	175.5	149.0	81.3	109.7	140.4	2.49	B
Potassium hydrogen bisphenyl acetate	177.7	180.0	177.7	113.9	113.9	2.44	A
Sodium hydrogen diacetate	166.8	137.7	166.8	110.8	110.8	2.45	A
Imidazolium hydrogen maleate*	49.8	38.7	49.6	109.6	109.2	2.37	A
SGPA	168.2	130.9	54.9	118.5	116.1	2.51	B
SGPB	173.0	159.5	173.0	112.5	112.5	2.49	A
Penicillopepsin Asp 33-Asp 213	123.8	149.8	147.1	159.0	147.4	2.88	B
Penicillopepsin Glu 16-Asp 115	168.4	134.1	163.7	123.1	113.0	2.75	B
Actinidin†	168.6	174.8	168.0	124.5	119.5	2.51	B

The angles are defined with respect to the carboxyl groups assigned, for type B acid salts, such that subscript numbers 1-4 refer to the $-\text{COOH}$ and, for proteins, to the group of lower sequence number.



Torsional angles: κ_1 , $\text{C}_1\text{C}_2-\text{O}_4\text{O}_5$; κ_2 , $\text{C}_2\text{O}_4-\text{O}_5\text{C}_7$; κ_3 , $\text{O}_4\text{O}_5-\text{C}_7\text{C}_8$. Bond angles: κ_4 , $\text{C}_2\text{O}_4\text{O}_5$; κ_5 , $\text{O}_4\text{O}_5\text{C}_7$. Distance: d , O_4-O_5 . Angles are measured in degrees. Standard deviations for the small molecule structures are of the order of 0.01 Å for bond length, 0.1° for bond angle and 0.2° for torsion angle. For the protein structures, the corresponding values are ~ 0.1 Å, 3.0° and 5.5°. Type A salts exhibit very short, apparently symmetrical hydrogen bonds typified by ^{22}OH stretching frequency in the 700-1,000 cm^{-1} region and bond strength of ~ 8 kcal mol^{-1} . * Ref. 21. † Ref. 16.

the proton by raising the pH leads to repulsion of the two negative charges, which in turn destabilizes the crystals and at pH values >4.5 , both SGPB and SGPA crystals dissolve. In a paper describing the refinement of SGPA, Seilecki *et al.*¹¹ present the interaction between the carboxyl group of Asp 123 and the terminal carboxyl group of Leu 242. This hydrogen-bonded interaction is of type B and includes a charge neutralizing Na^+ . The two carboxyl groups arise from separate molecules and thus contribute to the stability of the crystal structure. Parameters for this interaction are also included in Table 1.

The crystal structure of the aspartyl protease, penicillopepsin¹², has recently been refined at 1.8 Å resolution to an R factor of 13.6% (to be published). The carboxyl groups of the two catalytic aspartyl residues, Asp 33(32) and Asp 213(215), also share a proton at pH 4.4, the pH of the native crystals. These carboxyl groups are at the junction of the two large domains of the aspartyl proteases, at the bottom of the substrate-binding cleft. Their spatial arrangement is different from the examples given above (see Fig. 1b). As in SGPB, there is no balancing cation for the type B acid salt in penicillopepsin. However, the excess negative charge is in part balanced by the proximity of several peptide dipoles¹³. The carboxyl groups are also arranged so as to provide an electrophilic proton as well as to help stabilize the inter-domain junction. Furthermore, pepsin is reported¹⁴ to have a slightly raised pK_a for Asp 215(213) which is as expected where two carboxyl groups are maintained in close contact. That such a juxtaposition can cause an unfavourable interaction is seen in the structure of *Rhizopus chinensis* pepsin, where the homologous aspartyl residues adopt a totally different disposition to relieve the repulsion evident at pH 6.5 (ref. 15).

A further example of this is found in penicillopepsin (Fig. 1c), where the juxtaposition of the carboxylate of Asp 115(114) with that of Glu 16(13) demands that they share a proton, but the torsion angles (see Table 1) in this case seem more in keeping with a type A structure, although the small molecule data may not be representative. This seems to be associated with the additional presence of a hydrogen-bonding amide group from Asn 118(117) which donates the proton from Asn 118 N^{H2} to the other oxygen O^{e1} of Glu 16 ($d = 2.91$ Å). It can be seen that the carboxyl group of Glu 15 is not involved directly with these interactions.

The values of torsion angles $\kappa_1-\kappa_3$, bond angles κ_4 and κ_5 , and bond length d , for the proteins examined are not unusual when compared with those of the acid salts (Table 1). However, the values of κ_4 , κ_5 for the Asp 33(32)-Asp 213(215) interaction in penicillopepsin are appreciably higher and the distance separating the oxygen atoms is slightly longer than in the other examples. This is consistent with the less than ideal arrangement of proton sharing expected if the proton is to take part in catalysis.

The imposition of a restricting, near-neighbour arrangement

for two carboxyl groups can lead to an anomalous pK_a . Thus, the *cis*-double bond of maleic acid leads to the high pK_{a2} of 6.1 as compared to that of 4.4 for the *trans*-isomer, fumaric acid. In proteins it seems unlikely that an anomalous pK_a may arise from an analogous restrained environment around two carboxyl groups. Unfortunately the unambiguous assignment of anomalous titration data to particular carboxyls in proteins does not coincide with the highly refined structures available. For example, the actinidin structure at pH 5.9 (ref. 16) shows a type B acid salt interaction between glutamic acid residues 50 and 86 but no titration data are available. In the case of tobacco mosaic virus there is ample evidence for anomalous carboxyl groups^{17,18} but their interactions, although almost certainly of the type we describe, are not yet known in sufficient detail^{19,20}.

We may conclude that the carboxyl-carboxylate interaction is an important stabilizing force in protein-protein interactions at low pH and, at higher values, may occur more frequently than previously believed in association with an anomalous pK_a . Conversely, the unambiguous detection of an anomalous pK_a for acidic groups in proteins can be, and often is, a sign of carboxyl-carboxylate interactions of the type described here.

We thank W. Anderson for helpful criticism and C. Broughton, C. McPhalen and A. Sielecki for valuable assistance. L.S. thanks the MRC (Canada) for a Visiting Scientist Award and Lothian Regional Council for leave of absence. This work was supported by a MRC (Canada) grant to the Group on Protein Structure and Function, University of Alberta.

Received 30 June; accepted 26 October 1981.

- Pauling, L., Corey, R. B. & Branson, H. R. *Proc. natn. Acad. Sci. U.S.A.* **37**, 205-211 (1951).
- Speakman, J. C. *Structure & Bonding* **12**, 141-199 (1972).
- Golić, L. & Speakman, J. C. *J. chem. Soc.*, 2521-2530 (1965).
- Shrivastava, H. N. & Speakman, J. C. *J. chem. Soc.*, 1151-1163 (1961).
- Manojlović, L. & Speakman, J. C. *Acta crystallogr. B24*, 323-325 (1968).
- Speakman, J. C. & Mills, H. H. *J. chem. Soc.*, 1164-1175 (1961).
- Delbaere, L. T. J., Hutcheon, W. L. B., James, M. N. G. & Thiesen, W. E. *Nature* **257**, 758-763 (1975).
- Lundgren, L.-O. & Olovsson, I. in *The Hydrogen Bond* Vol. 2 (eds Schuster, P., Zundel, G. & Sandorfy, C.) 471-526 (North-Holland, Amsterdam, 1976).
- Hartley, B. S. *Nature* **201**, 1284-1287 (1964).
- Tang, J. *et al.*, *Proc. natn. Acad. Sci. U.S.A.* **70**, 3437-3439 (1973).
- Sielecki, A. R. *et al.*, *J. molec. Biol.* **134**, 781-804 (1979).
- Hsu, I.-N., Delbaere, L. T. J., James, M. N. G. & Hofmann, T. *Nature* **266**, 140-145 (1977).
- James, M. N. G., Hsu, I.-N., Hofmann, T. & Sielecki, A. R. in *Structural Studies on Molecules of Biological Interest* (eds Dodson, G., Glusker, J. P. & Sayre, D.) 350-389 (Oxford University Press, Oxford, 1981).
- Cornish-Bowden, A. J. & Knowles, J. R. *Biochem. J.* **113**, 353-362 (1969).
- Subramanian, E., Liu, M., Swan, I. D. A. & Davies, D. R. in *Acid Proteases Structure, Function and Biology* (ed. Tang, J.) 33-60 (Plenum, New York, 1977).
- Baker, E. N. *J. molec. Biol.* **141**, 441-484 (1980).
- Caspar, D. L. D. *Adv. Protein Chem.* **18**, 37-121 (1963).
- Durham, A. C. H. & Kung, A. *Nature new Biol.* **229**, 42-46 (1971).
- Stubbs, G., Warren, S. & Holmes, K. C. *Nature* **267**, 216-221 (1977).
- Bloomer, A. C., Champness, J. C., Brice, G., Staden, R. & Klug, A. *Nature* **276**, 362-368 (1978).
- James, M. N. G. & Matsushima, M. *Acta crystallogr. B32*, 1708-1713 (1976).
- Novak, A. *Structure Bonding* **18**, 177-216 (1978).

BOOK REVIEWS

Nationalized science and technology

Eric Ashby

THIS is a sober and sobering book. Professor Tisdell was asked "to prepare an independent background report on the subject of priority assessment in science and technology for the Australian Science and Technology Council". His book is an expansion of this report. Don't assume that it is of interest only to Australians; the prospects for Britain might be less depressing if a number of senior civil servants and members of the present government were to be given three days' leave in order to read the book and ponder its message.

When Queen Victoria came to the throne the pursuit of science, like the pursuit of literature, was a career for the self-employed, often subsidized by payment from some other occupation. T.H. Huxley earned his living at first by teaching; Anthony Trollope by inspecting post-offices. The notion that there was need for a science policy would have been as crazy as the notion that there was a need for a policy among poets and novelists.

By the time Queen Elizabeth came to the throne all this had changed. Poets and novelists could still pursue their stars without deviations caused by governments, but scientists were no longer immune from influence. The critical time was the year when the government grant to universities exceeded 50 per cent of their total income. There was no *dirigisme* at first, of course. The taxpayer had become paymaster and patron of science but initiative was left, still, to individuals. But the seeds of *dirigisme* had been sown. We are now painfully reaping the harvest: a rank growth of policy-documents, technology assessment, predictions about research and development. The public interest is now called a "customer" and the laboratory worker a "contractor": science by mail order.

This is, of course, the inevitable result of the staggering successes of science applied to human needs (though in one sense it would be more accurate to say science applied to human destruction). Governments have realized that research and development is too important to be left to industry alone. More and more scientific research — both pure and applied — is now the "property" of the taxpayer, whose money pays for government research establishments, the grants from research councils and a great deal of the costs of universities. The scientist, provided he is willing to work on problems for which he

Science and Technology Policy: Priorities of Governments. By C.A. Tisdell. Pp.222. ISBN 0-412-23320-7. (Chapman and Hall: 1981.) £12, \$25.

can get grants, can adapt himself to this situation. It is the customer who is in a dilemma. He has to draw up a shopping list and he hasn't enough money to pay for all the science he would like to have. So he has to decide priorities. And since "he", in a pluralistic democracy, is the social welfare function — the consensus of society — the task of deciding priorities is, on rational criteria, insuperable. The best one can do is to make provisional choices on imperfect evidence.

Tisdell's book is a manual of advice for those who make such choices. His starting points are the lists of research and development objectives published by OECD and the EEC. They are taxonomic lists of no great value. He then discusses an issue of special relevance to countries with small economies (compared with the giants), namely how much science should such countries import and exploit? And to what extent should such countries rely on their own springs of initiative in pure research? There is, he explains, a trade-off between the import of ideas and their local production. The profit to be made out of the exploitation of a new idea will vanish as soon as some (so called) less developed country starts marketing it at a fraction of the manufacturing cost.

The bulk of Tisdell's book is taken up with an analysis, well documented and supported by quantitative data, of the research and development policies of large economies (Germany, Japan, Britain and the USA) and small economies in the OECD (Belgium, Canada, Netherlands, Sweden and Switzerland). In all of these countries there is some formal procedure for setting priorities, sometimes well co-ordinated, sometimes (notably in Britain) fragmented among different government departments. For a while these procedures were conducted by bureaucrats and experts without reference to the public and without canvassing public opinion. This, as Tisdell demonstrates, is no longer practicable or desirable. Not only does the public want to participate through public interest groups — that is proving difficult enough — but also there are surges of pressure to change the order of priorities. Thus in the 1950s the assumption was to use profit and market success as the units for assessing research

and development. Then in the 1970s the assumption was challenged; improvement of the quality of life rose higher in the shopping list. Now we are back again to less exalted objectives: in a world depression, the high priority is to survive economically by capturing markets, and this has come to be the short-term aim of research and development.

In a sombre last chapter, Tisdell asks the questions — though of course he cannot supply the answers — that spin round in the policy-maker's mind all the time. Everyone admits that the optimum social welfare function cannot be derived from individual welfare functions; no one has a *politically* practicable alternative except to bring to politics the techniques of the oriental marketplace: haggling, trade-off and bluff. Not everyone, but many thoughtful people, admit that high unemployment among untrained people is a consequence of much modern technology; no one has a remedy for this evil except to repeat platitudes about making education more relevant (hardly a practicable proposal for schools on their present starvation diet).

At one point Tisdell suggests three options open to governments for science and technology: to try to slow down the rate of change; to let change proceed unplanned, in which individual governments and industries pursue their own self interest "with scant regard to the common good"; to establish collective or social priorities (and, I would add, international priorities) for the programme of change in science and technology. The first of these options is as feasible as stemming the tide; the second is an open cheque for chaos; the third is the only tolerable option. But will a nation as pluralistic as the United States, or as socially polarized as Britain is these days, adopt this option? President Carter's call for austerity to cope with "the moral equivalent of war" didn't meet with much response; Britain's attitude to the Brandt Report hasn't provoked a surge of altruism. The question, often asked and not yet answered, is whether human communities can make painful adaptations to avoid a breakdown in economic order before it comes; or whether, like communities of animals and plants, the adaptations will follow only as a response to a harsh environment after its effects have begun to bite. It is a sign of his anxiety that Tisdell turns at the end of his book to Daniel Bell's ominous book on the post-industrial society, with its "lack of an ordering

mechanism to make social choices". Tisdell warns: "We may all know that collectively we are heading for tragedy but individual self-interest of nations and of individuals may necessitate the tragedy to be played out". In a sensational pamphlet on doom such a sentence as this might be shrugged off. In an economist's sober report to the Australian Science and Technology Council, the sentence has a

ring of integrity and conviction; for Tisdell offers an antidote to despair: "A fall in exports and in living standards *could* be worthwhile from the point of view of achieving more basic ends". □

Lord Ashby was formerly Master of Clare College, Cambridge. His most recent book, written with Mary Anderson, is The Politics of Clean Air (Clarendon/Oxford University Press, 1981.)

From myth to methodology

C.K. Brain

Bones: Ancient Men and Modern Myths. By Lewis R. Binford. Pp.320. ISBN 0-12-100035-4. (Academic: 1981.) £24.20, \$36.50.

It is perhaps surprising that a new book by one of the most influential of archaeologists should deal, not with cultural remains, but with bones. By way of explanation, Lewis Binford observes that almost all of our ideas on the behaviour of prehistoric men have been based on interpretations of faunal remains and their depositional context — not, as so many textbooks would lead one to believe, of stone tools. So, in discussing reconstructions based on bones, Binford has significant things to say about the proper conduct of scientific enquiry as well as about the behaviour of early man.

Most of the inferences that have been drawn about the nature and activities of our prehistoric ancestors are seen by Binford as little more than "myths of prominent men". Some of these concepts may well have been accurate, but unless they were the products of a robust and reliable methodology for giving meaning to bones, they remain myths based on the opinions and perhaps even the biases of the investigating archaeologists. In traditional archaeology, Binford believes that concepts about early man have often been judged in accordance with the status of the archaeologist who propounded them. "We desperately need to abandon the technique of evaluating men and adopt the strategy of evaluating ideas", he writes, for in the absence of methodology it is people's reputations that are reshuffled.

The kind of methodology proposed by Binford as an alternative to myth-making is embodied in what he calls "middle-range research". A crucial characteristic of theories resulting from such research is that they are intellectually independent of the arguments about the phenomena they are used to illuminate. As an example, Binford cites the case of carbon-14 analysis, where independent physical research had established a reliable body of theory about the relationships between living organisms, atmospheric ^{14}C and rates of degeneration of the radioactive isotope. Such relation-

ships have been put to good use by archaeologists wishing to date organic remains preserved at their sites.

Apart from philosophical considerations of myths and methodologies, the main aim of the book is to provide a set of objective criteria for the recognition of human influence in the accumulation of a bone assemblage. What are the characteristic patterns of bone modification produced by non-human agents and how can these be distinguished from human modes of bone alteration? In attempting to answer these questions, Binford has drawn on his own extensive observations of Nunamiut Eskimos and their dogs in Alaska, as well as of wolves that inhabit those parts. He has provided a wide-ranging review of relevant work done elsewhere and has set up a convincing array of diagnostic criteria for separating hominid- from non-hominid-built bone accumulations.

Finally, in the third part of the book, Binford applies his methodology to bone assemblages from Olduvai Gorge, making use of data provided by Mary Leakey in Vol. 3 of *Olduvai Gorge*. As Clark Howell has pointed out in the foreword to Binford's book, these data were not collected for this specific purpose and so are not ideal. Nevertheless, some bold conclusions have been reached by Binford, such as that there is no evidence that Olduvai hominids hunted, or that they made use of base camps to which food was carried to be shared. Instead, our hominid ancestors are seen as having been scavengers at discarded carnivore kills where hammerstones were used to break open leg bones to extract marrow.

Binford's book is a major, if controversial, contribution to the rapidly expanding literature on the new science of taphonomy. Already his conclusions are being challenged, for example by H.T. Bunn (*Nature* 291, 574-577), and it will be interesting to follow the fortunes of his ideas under the impact of the rigorous methodology that he so eloquently advocates. □

C.K. Brain is Director of the Transvaal Museum, Pretoria, South Africa.

Nature's plumbing

Philip England

Geysers and Geothermal Energy. By John S. Rinehart. Pp. 223. ISBN 3-540-90489-1. (Springer-Verlag: 1981.) \$22.50, DM 38.
Geothermal Resources: An Energy Alternative. By Harsh K. Gupta. Pp. 227. ISBN 0-444-41865-2. (Elsevier Scientific: 1981.) \$61, Dfl. 125.
Geothermal Systems: Principles and Case Histories. Edited by L. Rybach and L.J.P. Muffler. Pp. 359. ISBN 0-471-27811-4. (Wiley: 1981.) £22, \$52.25.
Geothermal Systems. By John Elder. Pp. 508. ISBN 0-12-236450-3. (Academic: 1981.) £20.60, \$49.50.

STATED simply, a geyser is a water system to which mass and/or heat are added until a portion of the column is super-critical. At some stage nucleation of boiling occurs in this region and the consequent expansion partially empties the water column, reducing the pressure within it and leading to further boiling in the decompressed system, with spectacular eruptions of steam and water at the surface. The diversity of style and periodicity of geyser eruptions results from the almost endless variations on this theme which are possible given the multiple recharge-discharge systems occurring in nature.

In *Geysers and Geothermal Energy* John Rinehart has provided a detailed — one might say an affectionate — catalogue of these idiosyncrasies, together with the complex hydrological interpretations of them, when available. He also discusses briefly the geochemical and environmental aspects of geysers. One disappointing aspect of the book is that not enough attention is paid to the central process of geyser eruption: the problems of boiling initiation and bubble nucleation are dealt with in a cursory fashion. In other respects the book provides all the information on geysers that the layman or student could reasonably ask for. There is, in addition, one chapter of 15 pages on geothermal energy.

The remaining three books reviewed here all approach geothermal systems by way of an outline of plate tectonic theory, mantle convection and the present distribution of tectonic activity, and then diverge along lines that are to be expected from their formats. Gupta's *Geothermal Resources* is one of Elsevier's series on Developments in Economic Geology, and attempts to cover every aspect of the subject at the expense of the brevity which a 200-page format imposes on such an approach. The book is a review of the subject, covering everything from Earth structure and the basics of heat transfer (although convection is treated solely in terms of Newton's law of cooling) to production technology and world energy problems. The level and quality, however, are considerably lower than either of the two remaining books. The section on

global dynamics is not good, and nothing is dealt with in adequate detail. Further, the material is not particularly up-to-date and for a book of its price and brevity it has been poorly edited.

Geothermal Systems: Principles and Case Histories, a collection of papers edited by Rybach and Muffler, concentrates on specific topics. Following an introduction on geothermal anomalies and global tectonics, there are contributions on heat and mass transfer in hydrothermal systems, geothermal prospecting, heat extraction from reservoirs and resource assessment and environmental aspects, on which lucid and informative reviews (with up-to-date references) have been provided by people active in these particular fields. The reviews fill two-thirds of the book, the remainder consisting of case histories from various geothermal fields. These latter papers provide roughly the same sequence of information — geological framework, history of exploration, present-day exploitation — but are of a more variable quality than the first part of the book. The areas covered are the low enthalpy resource of the Pannonian Basin, the use of steam for electrical power generation from the Krafla field in Iceland, the Takinoue field in Japan and the Ahuachapan field in El Salvador, and the Jemez geothermal field — which contains high- and low-temperature hydrothermal systems as well as hot dry rocks. The book concentrates on the geological and geophysical aspects of geothermal energy extraction, so that there is little information on the engineering problems or the economics of this source of power.

To a greater extent than is usual for a collection of papers, this volume does succeed in giving a coherent picture of the subject. Credit for this must go to the editors who have chosen judiciously the areas to concentrate on and, I suspect, exerted some unifying influence on the contributions in the last one-third of the book.

If the antipodean reader were to feel that the geothermal fields of New Zealand are given unduly little attention in Rybach and Muffler's book, this hole is firmly plugged in Elder's volume of the same title. The book ought more properly to be called *Hydrothermal Systems*, as the possibility of extracting heat from dry rocks is not addressed. Instead, Elder provides a comprehensive survey of the mechanisms of heat and mass transfer within the Earth, at a higher level, but in a style which will be familiar to readers of the *Bowels of the Earth*. I am not referring to the use of anatomical analogies; that aspect is, mercifully, absent from the book (one may imagine the emunctory analogues which could be produced for hydrothermal systems).

The approach Elder adopts is more that of the engineer than of the applied mathematician or physicist; as he explains in the preface, this arises from a natural desire to

avoid deducing from first principles a set of results that are inconsistent with our observations of the Earth. It leads him to produce models for each system that he investigates which are characterized by pragmatism and simplicity, and which permit him to estimate order-of-magnitude figures for everything from global partial melting to the motion of steam bubbles in mud. Such an approach can be extremely successful (more so with bubbles than partial melt) but depends on the amount of common ground one shares with Professor Elder; certainly, it beats deriving every aspect of water-rock interaction from the Second Law of Thermodynamics, but it can be carried too far. After 400 pages of models being pulled out of whichever hat Professor Elder is wearing, either one suspends critical faculties and accepts each

model at face value (rather than working out the extent to which it matches geological reality), or one starts to feel the need for a more explicit physical base to what is being said. This is more of an error of judgement than a major drawback and there is much to gain from such a refreshing approach to the subject.

My only other criticism is of a similar order: the first 160 pages of this book are devoted to global tectonics and the origin and emplacement of magma bodies: aside from the fact that a newcomer to the subject would gain from it an impression that no geophysicist or geologist has worked on these problems, this is surely an over-long introduction. □

Philip England is in the Department of Geological Sciences, Harvard University.

Ecology: a Japanese perspective

John Lawton

Comparative Ecology, 2nd Edn. By Y. Itô. Edited by Jiro Kikkawa. Pp. 436. ISBN hbk 0-521-22977-4; ISBN pbk 0-521-29845-8. (Cambridge University Press: 1981.) Hbk £22.50, \$54; pbk £9.95, \$19.95.

SCIENCE, like pop music or *haute couture* has fads and cliques, although scientists like to pretend otherwise. Ecology is no exception, with too many groups isolated to a greater or lesser extent by language, geography, tradition and the impossibility of reading even a fraction of the relevant literature. Contact between European, North American and Japanese ecologists has generally been good, but inevitably many papers in Japanese go unread by English-speaking scientists. A major strength of Professor Itô's book is to summarize and publicize work originally published in Japanese — roughly ten per cent of the entries in the long and comprehensive bibliography are in this category.

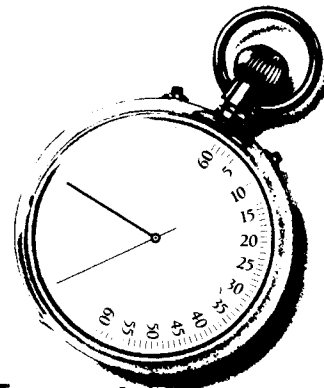
That said, I have to admit that I found the book heavy going, not because of Jiro Kikkawa's excellent translation, but because of the almost overwhelming mass of detail which Itô compresses between the covers, organized around a single unifying, but in the end not wholly convincing, theme.

For Itô, the animal kingdom is divided into beasts of the two sorts. On the one hand are those opting for high fecundity, living in environments where food for their young is easily procured; and on the other are species whose young enter a more hostile world, displaying low fecundity and parental care as a consequence. The whole of population dynamics, territorial behaviour and animal society are then interpreted within this simple framework.

Itô is at pains to show how his world view differs from the deeply entrenched traditions of "r and K selection" more familiar

to European and North American ecologists. Certainly there are differences, albeit rather subtle. More importantly, Itô's view stands or falls on his ability to provide an independent measure of food supply. On this fundamental point his thesis is distinctly shaky. Food supplies are never measured directly; indeed his logic is often

IRCS JOURNALS



Keeping pace with science

IRCS Medical Science
St. Leonard's House
St. Leonardgate,
Lancaster, U.K.

A member of Elsevier Science Publishers

Circle No. 12 on Reader Enquiry Card.

circular — if an animal displays parental care, its young, by definition, face an impoverished habitat.

To be fair, I suspect that Itô hopes to provoke others into making the necessary measurements, and like all attempts to force the bewildering variety of the living world into a manageable framework, his vision is certainly true in parts. Thus high, or low, fecundities (Chapter 1) do tend to be associated with particular types of survivorship curves (Chapter 2), generating particular patterns of population fluctuations (Chapter 3), and social systems (Chapters 4–6). But there are bound to be numerous exceptions, and in trying to force every example into one mould, the book sometimes makes difficult reading.

European and North American colleagues will also feel uneasy with those parts of the book which fail to incorporate

important concepts from their own contemporary literature. I noticed this particularly in Chapter 3, which deals with population fluctuations in a way that fails to reflect major advances of recent years.

I make these criticisms very reluctantly, because any book which increases the flow of ideas and information between groups of ecologists is to be welcomed. Nonetheless it must be said that this is not a good book to use with undergraduates, despite the fact that they may be more receptive to new ideas than teachers! But I do think that anybody interested in animal population dynamics and social behaviour should look at it, particularly for the richness and relative unfamiliarity of many of its principal characters. □

John Lawton is a Senior Lecturer in Ecology in the Department of Biology, University of York.

The inconstant biological membrane

A. Silberberg

Mechanics and Thermodynamics of Biomembranes. By E.A. Evans and R. Skalak. Pp.272. ISBN 0-8493-0127-0. (CRC Press/Blackwell Scientific: 1981.) \$34.95, £26.

MOST material systems are not mechanically simple. The stresses which arise cannot be expressed by the one- or two-parameter relationships which, nevertheless, have scored enormous practical successes. Much of fluid mechanics is based on the Navier–Stokes equations for an incompressible Newtonian fluid, and much of solid state deformation theory on a loss-less, purely elastic material constitutive equation.

The science of rheology has tended to bridge this gap in the engineering world, but in biology such complications are still mostly avoided. This is particularly reprehensible when, as is natural with living systems, movement of components, and indeed chemical reactions, have to be considered alongside any description of the stresses and strains which shape and confine the structural components. In developing the analysis of the chemical potential gradients along which transport of the various distinguishable chemical species will occur, the mechanical contribution is generally represented by a hydrostatic pressure gradient at most. Structural elements are taken out of consideration by being thought of as chemically inert and mechanically rigid.

This book by Evans and Skalak is thus a most welcome newcomer. It contains four main sections which, following an introduction, cover the kinematics, the dynamics and the thermodynamics of membranes; the final section summarizes the experimental situation. Some 100

references are given and a good index is also provided. The authors deal specifically with those membranes which confine cells or sub-cellular compartments, discussing the factors which mechanically govern the shape and thermodynamically will influence the stability of these systems.

This is, to my knowledge, the first systematic attempt to develop the considerations basic to the field in textbook form. The book is thus highly recommended to all workers in physiology and biophysics — to all scientists, in fact, who are dealing with the mechanics and thermodynamics of the interface in systems of heterogeneous and complex structure.

A. Silberberg is the Joseph and Marian Robbins Professor of Biorheology in the Department of Polymer Research, The Weizmann Institute, Rehovoth, Israel.

Electron microscopy

● A new volume in the series *Practical Methods in Electron Microscopy*, edited by Audrey Glauert, has recently been published by Elsevier/North-Holland Biomedical Press. Volume 9, *Dynamic Experiments in the Electron Microscope* by E.P. Butler and K.F. Hale, deals with the study of the response of materials to a change in state, using the transmission electron microscope, and costs hbk Dfl.195, \$95; pbk Dfl.92, \$45.

● M.A. Hayat's *Principles and Techniques of Electron Microscopy: Biological Applications* has now appeared in a second edition. Like its predecessor, the aim of the new edition is to provide a foundation in the biochemical concepts underlying the preparation of specimens for transmission electron microscopy. Co-published by University Park Press and Edward Arnold, the book costs \$34.50, £27.50.

MeV volume

M.H. Key

Laser Interaction and Related Plasma Phenomena, Vol.5. Edited by Helmut J. Schwarz *et al.* Pp.850. ISBN 0-306-40545-8. (Plenum: 1981.) \$75, £47.25.

LASER plasma interactions at high levels of irradiance involve a fascinating variety of physical processes and phenomena. Plasma produced in this way has many significant applications, arising mainly from the extreme energy density available. Laser fusion is the major motivation of much of the world-wide activity, but basic physics investigations — for example into X-ray lasers, energy transport in high temperature gradients and the characteristics of matter at extreme pressures — are beneficiaries of the technical developments arising from laser fusion research.

It is difficult to produce books which adequately describe the state of the art in such a wide-ranging and fast-moving field. The well-trying formula of a compilation of contributions from specialist authors has the best possibilities and the present volume is the latest in a series of this genre.

The old editorial team of H.J. Schwarz and H. Hora has been strengthened by M.J. Lubin and B. Yaakobi of the University of Rochester, and the book presents contributions from an internationally representative panel of 37 scientists who lectured at a workshop in 1979. I find that the book provides a better-organized selection of material than some of the earlier volumes; it has a mix of basic reviews and current research topics, and therefore offers something both for those researching in the same field and for those with a more general interest.

There are five major sections, covering lasers and alternative pulsed power sources, interaction experiments, general considerations for fusion, fusion-related experiments and interaction theory, and here the reader is given a fairly good guide to the topical issues of 1979–1980. The review material covers subjects such as new lasers, free electron lasers, light ion beams, new fusion fuels, high-density diagnostics, target manufacture, soliton theory, Rayleigh–Taylor instability, stimulated scattering and Stark broadening of high Z ion lines. It is generally authoritative and up to date. The descriptions of more detailed research activities have inevitably become somewhat dated, though specialists will find them useful as sources of reference to past work.

I feel researchers in the field will want access to this book as a valuable compendium of information, while non-specialists may wish to consult one or more of its review sections. □

M.H. Key is Deputy Head of the Laser Division at the SERC Central Laser Facility, Chilton, Oxfordshire.